GENOME WIDE IDENTIFICATION OF SELECTION SWEEPS IN HIGH ALTITUDE BOVINES FOR MILK PRODUCTION AND ADAPTABILITY TRAITS

THESIS SUBMITTED TO THE
ICAR-NATIONAL DAIRY RESEARCH INSTITUTE, KARNAL
(DEEMED UNIVERSITY)
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ANIMAL GENETICS AND BREEDING

BY

Dr. MOHAN M

(M. V. Sc)

DEPARTMENT OF ANIMAL GENETICS AND BREEDING
ICAR- NATIONAL DAIRY RESEARCH INSTITUTE
(DEEMED UNIVERSITY)
KARNAL, HARYANA 132001 INDIA

2022

Regd. No.: 18-P-AG-05

GENOME WIDE IDENTIFICATION OF SELECTION SWEEPS IN HIGH ALTITUDE BOVINES FOR MILK PRODUCTION AND ADAPTABILITY TRAITS

by
Dr. Mohan M
(M.V.Sc)

THESIS SUBMITTED TO THE
ICAR-NATIONAL DAIRY RESEARCH INSTITUTE, KARNAL
(DEEMED UNIVERSITY)
IN PARTIAL FULFILLMENT OF THE REQUIREMENT
FOR THE AWARD OF THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
ANIMAL GENETICS AND BREEDING

Approved by

MK Singh 17/06/2022
External examiner

17/6/2022
Major advisor
(Dr. S K Niranjan)

MEMBER OF ADVISORY COMMITTEE

Dr. Vikas Vohra
(Principal Scientist, Animal Genetics and Breeding, ICAR-NDRI)

Dr. Gopal Gowane
(Senior Scientist, Animal Genetics and Breeding, ICAR-NDRI)

Dr. Manishi Mukesh
(Principal Scientist, Animal Biotechnology, ICAR-NBAGR)

Dr. Sudarshan Kumar
(Scientist, Animal Biotechnology, ICAR-NDRI)

Dr. S S Lathwal
(Principal Scientist, Livestock Production and Management, ICAR-NDRI)
(Director's Nominee)

ICAR-NATIONAL DAIRY RESEARCH INSTITUTE

(DEEMED UNIVERSITY)

KARNAL - 132001, (Haryana), INDIA.

CERTIFICATE

This is to certify that the thesis entitled **"Genome wide identification of selection sweeps in high altitude bovines for milk production and adaptability traits"** submitted by **Dr. Mohan M** towards the partial fulfillment for the award of the degree of **Doctor of Philosophy** in **ANIMAL GENETICS AND BREEDING** of the **NATIONAL DAIRY RESEARCH INSTITUTE (Deemed University), Karnal (Haryana), India,** is a bonafide research work carried out by him under my supervision, and no part of the thesis has been submitted for any other degree or diploma.

Date: 21-06-2022

(Dr. S.K. Niranjan)
Major Advisor

Dedicated

to

The Almighty

My dear Parents

Teachers

&

Friends

ACKNOWLEDGEMENT

With the blessings of the "Almighty God", I completed this venture, I express my heartfelt sense of gratitude towards Omnipotent and all those who helped, guided, and inspired me to complete this study successfully.

I express my deepest gratitude and candid thanks to my esteemed advisor ***Dr S. K. Niranjan****, Principal Scientist, ICAR-NBAGR, Karnal for his inspiring and excellent guidance, encouragements, caring, timely suggestions, keen interest, cooperation and providing an excellent atmosphere during the entire research period without which it would not have been possible for me to complete this work. Its great honour to work with him.*

I am really grateful to the Director, ICAR-NDRI and Director, ICAR-NBAGR for their support. I also wish to thank the Head of Animal Genetics and Breeding Division, ***Dr Archana Verma****, for her never ending support. My knowledge I pledge to all my teachers who taught me in my course work. Thank you teachers.*

I would like to convey my regards and sincere thanks to the member of advisory committee ***Dr Vikas Vohra, Dr Manishi Mukesh, Dr Gopal Gowane, Dr Sudarshan Kumar, and Dr S S Lathwal*** *for the valuable suggestions, erudite guidance, and all possible cooperation during several phases of this investigation.*

A special thanks from my bottom of my heart and gratitude to ***Dr S Jayakumar****,* ***Dr Parameshwari*** *and* ***Dr K. N. Raja*** *for their technical assistance and moral support.*

I am nowhere without my Under Graduate teachers, ***Dr Palanisamy****, Dean, VC&RI, Tirunelveli,* ***Dr Jeyachitra****, HOD, Animal Genetics and Breeding, VC&RI, Orathanadu,* ***Dr Ashokan****, Former Dean, MVC and* ***Dr Gunaseelan****, Former Dean, VC&RI, Namakkal. I thank them so much for the care that they have showed. My sincere gratitude to my school teachers.*

I personally owe Dr Nandhini. P. B, Dr Ragulraj, Dr Praveen, Dr Dhandapani, Dr Arun, Dr Elango, Dr Joel, Mr Chandrasekar, Mr Alagu Niranjan, Dr Ashok kumar, Dr Utsav Surti, and Dr Balaganesh for their care which made me feel comfort

during this period. My seniors Dr Ravi, Dr Karthu for their love and brotherhood, thank you so much! Sweet thanks to Dr Surya, Dr Subrato Koloi, Dr Narendra pratap singh, Dr Sarin and Dr Destaw all my batch mates for all memorable moments.

All mess workers for their delicious food, Surender Bhaiya and Security guards of student's hostel. In all your memories.

I am whole heartedly thankful to my wife, ***Dr Sakthi Saranya****. K for bearing with all the hardships during the final days of my thesis. I would be nowhere without my family, my dad* ***Mr Malayaperumal****, my mom,* ***Mrs Banumathi****, my sister,* ***Mrs Malathi****, my brother-in-law,* ***Mr Ramki*** *and my God child,* ***Master Dheeshanth****, Love you guys.*

Date: ***MOHAN. M***

ABBREVIATIONS

A	:	Adenine
ARMS	:	Amplification Refractory Mutation System
AT	:	Adenine-Thymine
BAHS	:	Basic Animal Husbandry Statistics
BAM	:	Binary Alignment Format
BCF	:	Binary Calling Format
BED	:	Browser Extensible Data
BLAST	:	Basic Local Alignment Search Tool
bp	:	Base Pair
BTA	:	*Bos taurus* Autosome
C	:	Cytosine
CLR	:	Composite Likelihood Ratio test
CNV	:	Copy Number Variations
CRoPS	:	Complexity Reduction of Polymorphism Sequencing
DAHDF	:	Department of Animal Husbandry Dairying and Fisheries
ddRAD	:	Double digest RAD sequencing
DNA	:	DeoxyriboNucleic Acid
EDTA	:	Ethylene Diamine Tetra Acetic acid
FAO	:	Food and Agriculture Organization
FASTA	:	Fast all
FLK	:	Fixation by extension Lewontin and Krakauer statistics
Fst	:	Fixation Index
G	:	Guanine
Gb	:	Giga base pair / Giga bytes
GBS	:	Genotyping by Sequencing
gDNA	:	genomic DNA
GFF	:	General Feature Format
GoI	:	Government of India
GTF	:	General Transfer Format
GWAS	:	Genome Wide Association Studies
GWSS	:	Genome wide sampling sequencing

hapFLK	:	haplotype based FLK
HD	:	High Density
HIF	:	Hypoxia Inducible Factors:
HKA	:	Hudson Kreitman Aguade
iHS	:	integrated Haplotype Score
INDEL	:	Insertion and Deletion
Kb	:	Kilo base pair / Kilo bytes
KEGG	:	Kyoto Encyclopedia of Genes and Genomes
LD	:	Low Density/ Linkage Disequlibrium
Mb	:	Mega base pair / Mega bytes
MKT	:	McDonald Kreitman test
NCBI	:	National Center for Biotechnology Information
NGS	:	Next Generation Sequencing
OD	:	Optical Density
PCR	:	Polymerase Chain Reaction
pO2	:	Partial Pressure of Oxygen
QC	:	Quality Control
QTL	:	Quantitative Trait Locus
RAD	:	Restriction Site associated DNA Sequencing
RD	:	Read Depth
RE	:	Restriction Enzyme
rEHH	:	relative Extended Haplotype Homozygosity
RFLP	:	Restriction fragment length polymorphism
RNA	:	Ribonucleic Acid
ROH	:	Runs of Homozygosity
RPM	:	Revolutions per minute
rpm	:	Revolutions Per Minute
RRL	:	Reduced Representation Libraries
SAM	:	Sequence Alignment Format
SBH	:	Sequencing By Hybridization
SDS	:	Sodium Dodecyl Sulfate
SFS	:	Site Frequency Spectrum
SNP	:	Single Nucleotide Polymorphism
SOLiD	:	Sequencing by Oligonucleotide Ligation and Detection

SSCP	:	Single Strand Conformation Polymorphism
SweeD	:	Sweep Detector
T	:	Thymine
TEB	:	Tris base, boric acid and EDTA
Ti	:	Transition
TV	:	Transversion
UV	:	Ultraviolet
VCF	:	Variant Calling Format
WGS	:	Whole Genome Sequencing
XP-CLR	:	Cross-Population Composite Likelihood Ratio
XP-EHH	:	Cross Population Extended Haplotype Homozygosity

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Ladakh (UT), a unique habitat with an altitude of more than 3000 m above mean sea level is the home of unique farm bovine species, well adapted to hypoxic conditions, subnormal temperature, high UV exposure, and sparse feed resources. The unique genetic architect of these native bovines is an interesting venture to study the adaptation; it also needs to explore the genomic region of milk production for their further genetic improvement. Genome wide comparison of high producing Zebu and exotic cattle with high altitude bovines might help to explore variety of hypothesis that may be utilized for genetic improvement for milk of native bovines without changing the genetic architect for adaptation. With hypothesizing the uniqueness of high altitude bovines and their differences with dairy cattle breeds, a technical program has been proposed with study group comprises native cattle (12), yak (12) and their hybrids (6) of Ladakh, and high producing cattle Sahiwal (12), Holstein Friesian (12) and Jersey (12). The samples were sequenced using ddRAD technology, data analysis done by creating a bioinformatic pipeline from open source tools. A total of 610058, 357981 and 311100 SNPs with novel proportion of 59.89, 86.93 and 81.26 % were detected in Ladakhi cattle, yak, and its hybrids respectively. Genome wide SNP density analysis showed a SNP occurs at every 4336, 7448 and 8563 bp intervals in Ladakhi cattle, yak, and its hybrids respectively. Total 59 high SNP dense regions were found common in all three-group influencing vital traits related to production, reproduction, and adaptation traits. Gene and Phenotype ontology of missense variants bearing genes accounted for the structural and physiological changes in high altitude bovines. Selection sweep scanning revealed 131, 78 and 67 highly significant genes were selected in individually Ladakhi cattle, yak, and its hybrids. These genes are involved in adaptive interactive gene networks influencing oxygen carrying, cardiovascular, cutaneous and energy metabolism systems. Selection sweep comparison between high altitude bovines with high producing cattle, showed former has similar selected regions related to milk production with high producing cattle and later has similar selected regions related to reproduction and adaptation with high altitude cattle. Among the High producing cattle based on greater proportion of selected regions, ranked correspondingly as Sahiwal, followed by Holstein Friesian and Jersey. This study implies that genetic variants occurred in high altitude bovines could be a retroaction to local environment. Adding value to above identified selection sweep also gives an idea on the gene level adaptation aided by nature's selection process undergone in Himalayan bovines to cope up with harsh conditions and which would help for marker-based selection in future. Comparative selection sweep analysis of high-altitude bovines with high producing cattle throws light on the wide possibilities that could be used for selection and breeding strategies for breed improvement which in turn might develop socio-economic status of locals.

सार

लद्दाख (यूटी), समुद्र तल से 3000 मीटर से अधिक की ऊंचाई के साथ एक अद्वितीय आवास, अद्वितीय कृषि गोजातीय प्रजातियों का घर है, जो हाइपोक्सिक स्थितियों, असामान्य तापमान, उच्च यूवी जोखिम और दुर्लभ फ़ीड संसाधनों के अनुकूल है। इन देशी गोवंशों का अद्वितीय आनुवंशिक वास्तुकार अनुकूलन का अध्ययन करने के लिए एक दिलचस्प उद्यम है; इसके आगे आनुवंशिक सुधार के लिए दूध उत्पादन के जीनोमिक क्षेत्र का पता लगाने की भी आवश्यकता है। उच्च उत्पादक ज़ेबू और उच्च ऊंचाई वाले गोजातीय पशुओं की जीनोम व्यापक तुलना विभिन्न प्रकार की परिकल्पनाओं का पता लगाने में मदद कर सकती है जिनका उपयोग अनुकूलन के लिए आनुवंशिक वास्तुकार को बदले बिना देशी गायों के दूध के आनुवंशिक सुधार के लिए किया जा सकता है। इस विश्वास को पूरा करने के लिए, अध्ययन समूह के साथ एक तकनीकी कार्यक्रम प्रस्तावित किया गया है जिसमें लद्दाख के देशी मवेशी (12), याक (12) और उनके संकर (6) और उच्च उत्पादक मवेशी साहीवाल (12), होल्स्टीन फ्रेसियन (12) और जर्सी शामिल हैं। (12)। नमूनों को डीडीआरएडी तकनीक का उपयोग करके अनुक्रमित किया गया था, ओपन सोर्स टूल्स से जैव सूचनात्मक पाइपलाइन बनाकर डेटा विश्लेषण किया गया था। लद्दाखी मवेशियों, याक और इसके संकरों में क्रमशः 59.89, 86.93 और 81.26% के उपन्यास अनुपात के साथ कुल 610058, 357981 और 311100 एसएनपी पाए गए। जीनोम वाइड एसएनपी घनत्व विश्लेषण से पता चला है कि लद्दाखी मवेशियों, याक और इसके संकरों में क्रमशः प्रत्येक 4336, 7448 और 8563 बीपी अंतराल पर एक एसएनपी होता है। उत्पादन, प्रजनन और अनुकूलन लक्षणों से संबंधित महत्वपूर्ण लक्षणों को प्रभावित करने वाले तीनों समूहों में 59 उच्च एसएनपी घने क्षेत्र सामान्य पाए गए। उच्च ऊंचाई वाले गोजातीय में संरचनात्मक और शारीरिक परिवर्तनों के लिए जिम्मेदार जीन वाले मिसेंस वेरिएंट के जीन और फेनोटाइप ऑन्कोलॉजी। चयन स्वीप स्कैनिंग से पता चला कि 131, 78 और 67 अत्यधिक महत्वपूर्ण जीन व्यक्तिगत रूप से लद्दाखी मवेशियों, याक और इसके संकरों में चुने गए थे।ये जीन ऑक्सीजन ले जाने, हृदय, त्वचीय और ऊर्जा चयापचय प्रणालियों को प्रभावित करने वाले अनुकूली इंटरैक्टिव जीन नेटवर्क में शामिल हैं। उच्च उत्पादक मवेशियों के साथ उच्च ऊंचाई वाले गोजातीय के बीच चयन स्वीप तुलना से पता चलता है कि पूर्व में उच्च उत्पादन वाले मवेशियों के साथ दूध उत्पादन से संबंधित समान चयनित क्षेत्र हैं और बाद में उच्च ऊंचाई वाले मवेशियों के साथ प्रजनन और अनुकूलन से संबंधित समान चयनित क्षेत्र हैं। चयनित क्षेत्रों के अधिक अनुपात के आधार पर उच्च उत्पादन करने वाले मवेशियों में, साहीवाल के रूप में रैंक किया गया, उसके बाद होल्स्टीन फ्रेज़ियन और जर्सी का स्थान है। इस अध्ययन का तात्पर्य है कि उच्च ऊंचाई वाले गोजातीय में होने वाले आनुवंशिक परिवर्तन स्थानीय पर्यावरण के लिए एक प्रतिगामी हो सकते हैं। उपरोक्त पहचान किए गए चयन स्वीप में मूल्य जोड़ने से कठोर परिस्थितियों से निपटने के लिए हिमालयी गोजातीय में प्रकृति की चयन प्रक्रिया द्वारा सहायता प्राप्त जीन स्तर अनुकूलन पर एक विचार भी मिलता है और जो भविष्य में मार्कर आधारित चयन में मदद करेगा। उच्च उत्पादन वाले मवेशियों के साथ उच्च ऊंचाई वाले गोजातीय का तुलनात्मक चयन व्यापक विश्लेषण उन व्यापक संभावनाओं पर प्रकाश डालता है जिनका उपयोग नस्ल सुधार के लिए चयन और प्रजन रणनीतियों के लिए किया जा सकता है जो बदले में स्थानीय लोगों की सामाजिक-आर्थिक स्थिति विकसित कर सकते हैं।

CHAPTER –1

Introduction

INTRODUCTION

India has a vast geographical, climatic, and biological diversity, hence righteously referred as "Subcontinent". Having about 2.4 percent of global geographical region, the country possesses 7.6 percent of mammalian, 12.6 percent of avian, 11.7 percent of fish, and 6.0 percent of all flowering plant species available in the world. With 535.8 million of livestock and 851.8 million of poultry population, the country is also rich in its animal genetic resources for food and agriculture (DAHDF, 2019). India's livestock sector is globally one of the largest, comprising 57.6 percent buffaloes, 12.8 percent cattle, 21.4 percent small ruminants, 2.1 percent camel, 1.2 percent equine, 1.1 percent pigs and 3.6 percent poultry population of the world (BAHS, 2020). India has 199 livestock and poultry breeds and many specialized populations, which are distributed in diverse regions, landscapes, and climatic conditions. Some of the regions are very harsh and even adverse for survival, however, the native livestock of such a landscape still thrive normally and provide the animal food to the local society.

Ladakh (UT), a trans Himalayan region at 3,000 m above Mean Sea Level (MSL), known as "cold desert", is one of such difficult landscapes in India. The region is characterized by very low ambient temperature, ranging from -40 to 20 °C, low barometric pressure, high Ultraviolet rays' exposure, only 100 mm annual precipitation per annum, generally, in the form of snow and poor vegetation (Bharti et al., 2017). Since agricultural cropping is highly limited in the region due to unfavourable climatic conditions in most part of the year; the livelihood of the native people is largely dependent on the native livestock. The livestock in Ladakh has been found to contribute up to 86 percent of the total household (Akand et al., 2017). Total livestock population in Ladakh region is about 6.4 lakh, including 2.77 lakh (43.4 percent) domesticated bovines (Shergojry et al., 2017). The bovines, adapted to high altitude conditions include native Ladakhi cattle (*Bos indicus*), yak (*Bos grunniens*) and yak X cattle hybrids (Dzo-Dzomo), which are kept mainly for milk and. The Ladakhi cattle and yak, although reared at different altitudes, are also genetically related and share chromosomal analogy. The Ladakhi cattle are reared at about 3000-3500 m msl and produce 2- 4 kg milk / day; whereas the Ladakhi yaks are reared over 4000 m msl of alpine region and produce about 0.5 - 1.5 kg milk/day. The yak-cattle hybrids, preferred by a large section of the society,

are generally reared at mid altitude regions of Ladakh (3500 to 4000 m msl) and produce high quantity of milk (up to 5 kg / day) (Akand et al., 2017; Shergojry et al., 2017). The milk productivity of these bovines has been limited, possibly because of fewer genetic selection efforts. Further, adaptation through modified morphological appearance and feedback mechanism for balancing energy deficit to sustain at higher altitude also seemingly restricted their milk production ability. In recent times, to increase the productivity of the native bovines, a crossbreeding program of native cattle with exotic Jersey (*Bos taurus*) has been adopted by the local government agencies.

Survivability of human as well as animals is highly challenging in high altitude regions due to low temperature, sparse food resources, high level of radiation, lower oxygen partial pressure (pO_2) and negative energy balance (Lenfant, 1971). The livestock species of high altitude have developed numerous morphological and physiological adaptive traits, like thick hair fibres, double subcutaneous fat, large blood vessels, higher capacitance of lung and heart to withstand subnormal temperature and low oxygen content (Bouverot, 2012). Thus, nature prompts a strong selection pressure and acts on advantageous mutation to spread among population. Inevitably, the frequency of linked neutral variants also increases, called *genetic hitchhiking effect* forming a selective sweep region (Smith and Haigh., 1974; Nielsen et al., 2005; Sabeti et al., 2012).

Selective sweep regions tend to influence quantitative trait loci (QTL) present within those regions (Chevin., 2019). Many recent studies reported selective sweep regions associated with production, reproduction, and adaptation traits, notably milk yield, composition, fat yield and protein yield in dairy cattle (Cheruiyot et al., 2018; Aviles et al., 2013). Puglisi et al., (2013) found selected regions related to fertility which influences oocyte growth, puberty, and conception rate in livestock. Hypoxia Inducing Factor (HIF) pathways responsible for red blood cells production and decreasing blood flow resistance were influenced by selected regions reported by Lorenzo *et al*., (2014) and Witt and Sanchez., (2019) in high altitude conditions.

Recent advances in molecular technologies have helped in finding novel markers for the traits related to production, reproduction, and adaptiveness in livestock species. Various developments occurred in sequencing technology among those combining Reduced Representation Library (RRL) and DNA barcoding methods followed by high throughput sequencing (Elshire *et al.*, 2013). One of those RRL methodologies is double

digestion Restriction Site Associated DNA (ddRAD) sequencing, known for its site selection-based library, and supposed to be effective in finding out the novel SNPs without bias ascertainment. The ddRAD is gaining popularity because of its cost-effectiveness and easier computational pipelines, which made genome wide downstream analysis to identify positive selected signatures much possible (Zhou et al.,2014, Yang et al., 2016).

The genetic architecture of the high-altitude bovines is an interesting venture for exploring genome wide distribution of the selection sweeps particularly for milk and adaptation traits and their usefulness for increasing productivity in the native bovines of Ladakh. The regions for milk production are needed to be compared with high producing zebu and exotic cattle, so that the most complementary genotypes may be utilized for genetic improvement for the milk of native bovines without diluting the genetic architect associated with adaptation traits. By knowing the evolutionary relationship among the livestock species of Ladakh and comparing with the world's highest milk producing cattle breeds, identified genome-wide sweeps would be useful for prior selection of the animals/breed/species. Therefore, following objectives have been designed with the aim to explore the genetic architect and selection mechanism for milk and adaptation traits in high altitude bovines, using ddRAD sequencing.

1. **To identify genetic variants and selection sweeps in Ladakhi cattle, yak, and their hybrids**
2. **To functionally annotate genetic variants and selection sweeps for milk production and adaptation in high altitude bovines**
3. **To compare identified selection sweeps of high altitude adapted bovines with high producing *Bos taurus* and *Bos indicus* cattle**

CHAPTER –2

Review of Literature

REVIEW OF LITERATURE

2.1 Mammalian genome

The Mammalian genome comprises 3 billion base pairs of nucleotides with approximately 22 thousand genes, among which 14 thousand are found to be common to all mammalian species (Zimin *et al*., 2009). With the invention of DNA sequencing technologies researchers started exploring genetic makeup of organisms at nucleotide level and created the first ever reference genome in 2003 for humans (Rexroad et al., 2019). Reference genomes were used as a prime source for data mining of test populations and discovery of new vital novel mutations. Thus, genome research helped to better understand biology and evolutionary aspects of the species, local adaptation, disease resistance and more (Elisk *et al*., 2014). Livestock genomics started blooming after the constructing reference genome of *Bos taurus* cattle by Zimin et al., (2009) which opened numerous ways to figure out the economically important candidate genes, novel mutations and QTLs related to milk production that also aided in the selection process of the elite population. With similar goal research attempted to create a reference genome for numerous non model mammalian species such as Zebu (*Bos indicus*), Yak (*Bos grunniens*), Indian Mithun (*Bos frontalis*), and Gaur (*Bos gaurus*) (Canvez et al., 2012; Zhang et al.,2021; Mukherjee et al., 2019; Hassanin et al.,2012) in the hopes of better understanding the genetic architecture using whole genome sequencing. Recent advances in genomic research, as well as a variety of downstream analyses, have led to a new venture of identifying genomic regions that are positively selected for a variety of characteristics as a form of local adaptation that can help us relate to a variety of environmental patterns and to selection (Taylor et al., 2016; Muchadeyi et al., 2020)

2.2 Genetic variants

Any change in the genome is called variants; however, it turns significant when it becomes heritable. Genetic variation refers to diversity in gene frequencies causing differences between individuals or populations. It occurs in an individual or population primarily by mutation, but mechanisms such as sexual reproduction and genetic drift also contribute to produce genetic variation (Ku et al., 2010). Based upon its nature of occurrence, these can be classified as Sequence variant (Single Nucleotide Polymorphism (SNPs) and Insertion-Deletion (Indels)) and Structural variants (Insertion,

Inversion, Deletion, Duplication, Copy Number Variation (CNVs). SNPs, Indels, and CNVs have been the focus of variant studies in recent years; among these, SNPs and Indels have gained importance due to their abundance in the genome (Table 2.1) (Cirulli and Goldstein., 2010).

Table 2.1: Types of variants and their distribution in *Bos taurus* genome

Variant type	Rare (<1%)	Common (>1%)
SNP	150,000	3,000,000
Indel	15,000	250,000
CNV	150	2000

2.2.1 Single nucleotide polymorphisms

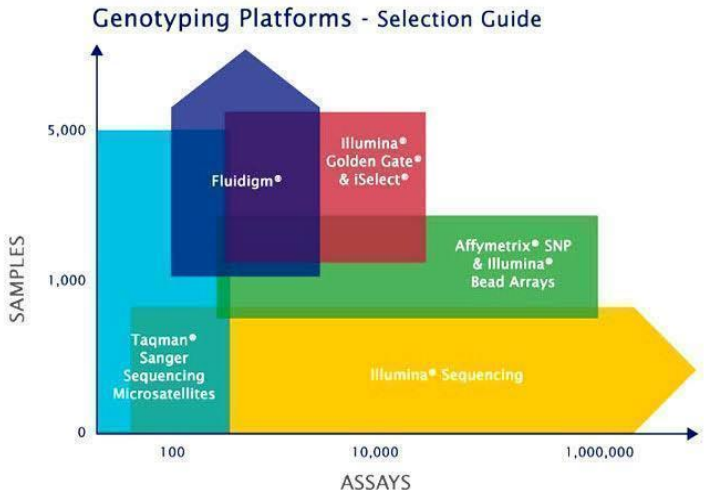
Single Nucleotide Polymorphisms (SNPs) are variations in DNA sequence that occur when a single nucleotide i.e., Adenine (A), Thymine (T), Cytosine (C) or Guanine (G) in the genome sequence gets changed. Numerous studies have discovered more than 100 million SNPs in various populations throughout the world, and these variations may be unique or exist in many individuals (Zwane et al., 2019). SNPs are more trustworthy than other markers because they are abundant and densely dispersed across the genome (434 bp in Bos taurus, 104 bp in Bos indicus) and have a low mutation rate ($1X10^{-9}$) despite their biallelic character. As a biological marker, it aids in the discovery of genes linked to economically significant quantitative and qualitative features (Yang et al., 2013; Donato et al., 2013). Any DNA that is lost (deletion) or gained (insertion) on a smaller scale is referred to as an indels. Indels are classed as Frameshift and in-frame if they occur in the coding region. Frameshift indels are more likely than in-frame indels to result in premature stop codons and have a greater functional impact (Sanders and Mason, 2016)

2.2.2 SNPs genotyping techniques

There are two basic ways for genotyping SNPs: traditional and high throughput approaches. Traditional gel-based SNPs genotype methods such as an amplification refractory mutation system (ARMS), restriction digests, and various types of gel electrophoresis (e.g., RFLP), denaturing gradient gel electrophoresis (DGGE), and

single-strand conformation polymorphism (SSCP). These methods are allele specific designed to find few numbers of SNP markers and has disadvantages of time consuming and low genome coverage (Koopaee et al., 2014). Hence traditional gel based SNPs genotype methods were slowly replaced by a variety of modern sequence based approaches including Microarray and Next Generation Sequencing (NGS) platforms, which has the advantage of identifying large numbers of variants and could have greater coverage depth. (Figure 2.1).

Figure 2.1: Various SNP-genotyping platforms



(Source: Unterseer et al., 2014)

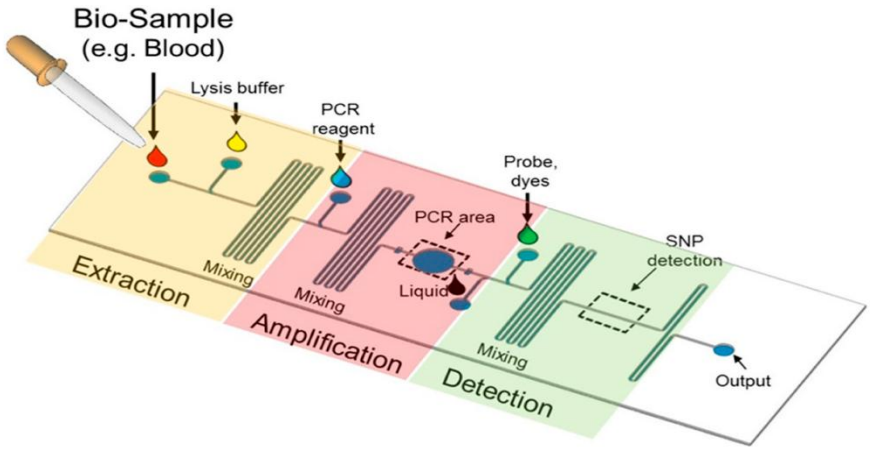
In recent times, many high throughput platforms like allele discriminationmethods (Allele-Specific Hybridization, Allele-Specific Single-Base Primer Extension), High-throughput assay chemistry (Flap endonuclease discrimination, Oligonucleotide ligation), DNA arrays, pyrosequencing and light cycler are available for SNPs genotyping that has a unique combination of coverage scale, accuracy, throughput, and cost (Kim and Misra., 2007; Gabriel *et al*. 2009). Considering application in livestock's improvement most studies were based on DNA arrays and High throughput sequencing.

2.2.3 DNA arrays

DNA chips and microarrays are immobilized oligonucleotides of known sequences, which differ at specific sites of individual nucleotides (at the site of SNP) and can also be used for the detection of known SNPs. It makes use of the technique of sequencing by hybridization (SBH) and involves tiling strategy. Four oligonucleotides in a column of an array differ only at the SNP site and only one would be fully

homologous. When such an array is hybridized with the PCR product, the perfect match allows the binding and mismatched products would be washed away. The perfect match in each case can be detected through a detection system (Huang et al.,2015; Nicolazzi et al., 2015). Various SNP chips were developed for major livestock species over time for fast detection of known valuables and to assist in the selective breeding process. Commercially available SNP chips are listed in Table 2.2.

Figure 2.2: Protocol for DNA SNP chip/ Microarray



(Source: Huang et al., 2015)

Table 2.2: SNPs chips developed for various livestock species

Species	SNP Chips	No. of SNPs	Developer
Cattle	3K	2900	Illumina
	LD v. 1	6909	Illumina
	LD v. 1.1	6912	Illumina
	GGPLD v1	8610	Neogen Geneseek
	GGPLD v2	19721	Neogen Geneseek
	GGPLD v3	26151	Neogen Geneseek
	SNP50 v.1	54001	Illumina
	SNP50 v.2	54609	Illumina
	GGPHD	76879	Neogen Geneseek
	AxiomBos1	648875	Affymetrix
	HD	777962	Illumina
	INDUSCHIP (Bos indicus)	43516	NDDB-India

Buffalo	Axiom Buffalo genotyping array BUFFCHIP	90000 57000	Affymetrix NDDB and USDA
Sheep	SNP50 v.1 HD	54241 606006	Illumina Illumina
Goat	SNP50 v.1	53347	Illumina
Pig	GGPLD v.1 SNP60 v.1 SNP60 v.2 SNP80	10241 62163 61565 68528	Neogen Geneseek Illumina Illumina Neogen Geneseek
Horse	SNP50 v.1 SNP70	54602 65157	Illumina Neogen Geneseek
Chicken	Axiom Chicken	580961	Affymetrix

2.2.4 Next Generation Sequencing platform

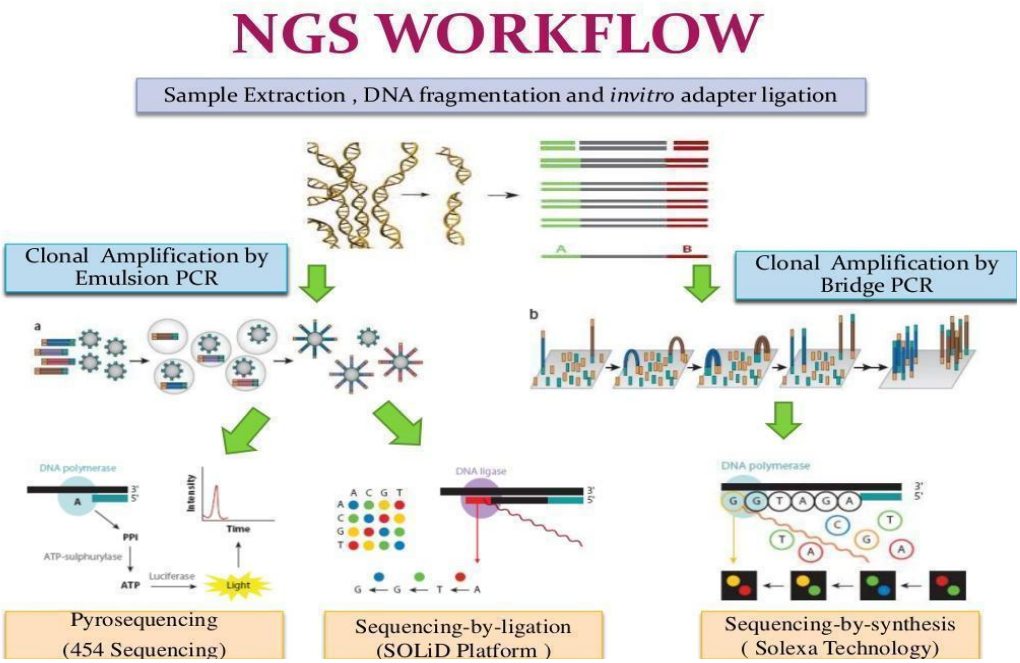
Till the 21st century Maxam Gilbert and Sanger methods of sequencing dominated omics related research, but later introduction of Next Generation Sequencing (NGS), became more easy, efficient, and cost-effective. DNA sequencing by NGS technologies has made a revolutionary impact on genome wide analysis finding mutation, association, and diversity (Delseny *et al*, 2010). Newer NGS technologies constitute various strategies based on preparation of template, sequencing and imaging and alignment of genome assembly methods (Metzker, 2010). Presently, there are various NGS platforms *viz.*, GS FLX, Roche 454, Illumina, Solexa, SoLiD, Ion Torrent, PacBio differing in their principal and chemistry are available (Table 2.3) in the market (Hamilton and Buell, 2012). With this technology the sequence reads are obtained from the fragmented DNA libraries, and there is no vector-based library preparation. Even though the sizes of the NGS reads are much shorter (25-250 bp) than that of capillary sequencing (650-800 bp), the parallel processing of millions of reads provides very high coverage of the genome (Mardis, 2008). The NGS technology has various advantages over SNP chip arrays in relation to application such as it helps to find novel SNPs and can detect other variants such as SSRs and InDels without any bias (Roh et al., 2010)

Table 2.3: Comparison of DNA sequencing platforms

Attributes	Roche 454/GS FLX +	Illumina GAII		Life Technologies / SOLiD
		GAII	HiSeq 2000	
Library	Fragment/emulsion PCR fragment	colony fragment		emulsion PCR
Sequencing	Pyro-sequencing	Sequencing by synthesis		Sequencing by ligation
Read length (base)	700–1000	150	100	75
Gb per run	0.7	95	600	300
Pros	Improve mapping, fast run time	most widely used		Error correction
Cons	High cost, high error rate	Low multiplexing capability of Samples		Long run time

(Source: Liu et al, 2011)

Figure 2.3: Overview of NGS work-flow



(Source: Yin et al., 2016**)**

2.2.5 Restriction site associated DNA (RAD) sequencing

NGS based Restriction site associated DNA sequencing (RAD) sequencing methods are an alternative to WGS, wherein simultaneous sequencing, genotyping, and multiplexing are facilitated (Young et al., 2010; Luca et al.,2011). RAD Sequencing methods are also known as Genotyping by sequencing (GBS), Genome wide sampling sequencing (GWSS). This method is having advantage of minimizing the Repetitive sequences, and thereby reducing the cost of sequencing and genotyping. Thus, we can sequence thousands of individuals in a matter of weeks. There are various RAD sequencing methods Complexity Reduction of Polymorphism Sequencing (CRoPS), Reduced Representation Libraries (RRL), Restriction Site associated DNA Sequencing (RAD seq), Genotyping by Sequencing (GBS), Double-digest RAD sequencing (dd RAD) (Peterson *et al*., 2012; He et al., 2014), which are having their own advantages and disadvantages as described in Table 2.4.

Table 2.4: Comparison of various RAD sequencing methods

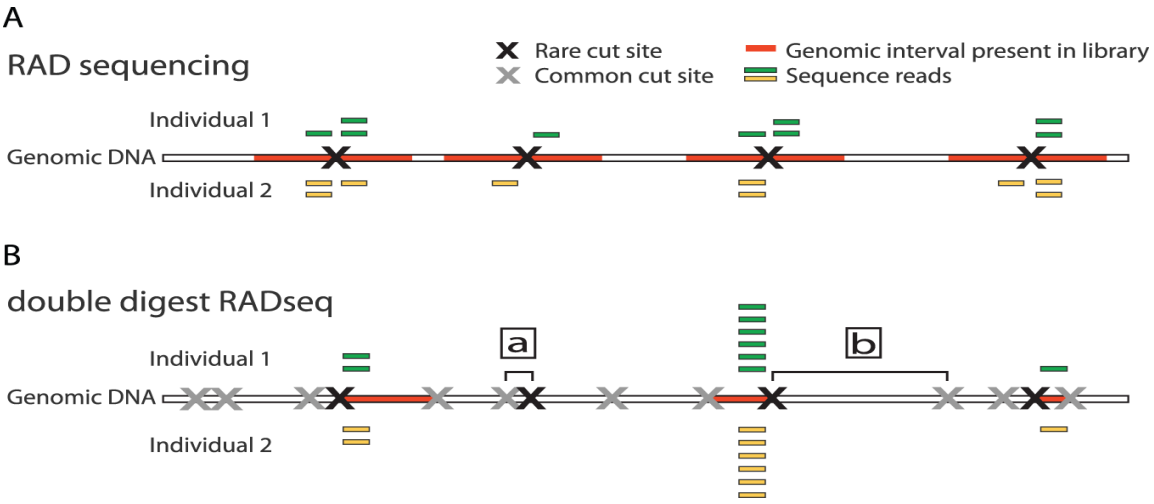
Methods	RAD	GBS	RRL	ddRAD
Digestion by RE	Yes	Yes	Yes	Yes
Ligation	Yes	Yes	Yes	Yes
Pooling	Yes	Yes	Yes	Yes
Shearing	Yes	No	No	No
Size selection	Yes	No	Yes	Yes
Multiplexing	Yes	No	No	Yes
Remarks	DNA loss	No Size Selection	No Multiplexing	Multiplexing

2.2.6 Double digestion restriction site associated DNA Sequencing (ddRAD seq)

In the Double digestion Restriction site Associated DNA Sequencing (ddRAD) sequencing, the fragment shearing is replaced by a second restriction digestion to increase the tunability and accuracy of the size-selection stage (Figure 2.4). In this process, the genomic DNA is first digested with a restriction enzyme and a barcoded P1 adapter linked to the fragments. Adaptor-linked fragments from different samples are combined, if the samples are multiplexed and the DNA is digested by a second

restriction enzyme. The size of the fragments is determined, and they are cleansed. These fragments are then amplified once the P2 primers are joined together (Peterson *et al*., 2012; Yang *et al*., 2016)

Figure 2.4 Overview of ddRAD library



Double digestion is followed by precise size selection that excludes regions flanked by either very close or very distant RE recognition sites, recovering a library consisting of only fragments close to the target size (red) (Source: Peterson et al., 2012)

2.3 Genome wide Selection Signatures:

Selection processes affect patterns of genetic variation in the neighbourhood of selected loci, resulting in variation of patterns in these regions and marks as a specific signature. The selective signatures are positive driven force which causes rise of certain individual fitness causing increase of between population and decreases within population variances. It may be classified as hard or soft based on the origin, type, and frequency of mutation. Hard selective sweep (classic sweep) causes the frequency of a rare beneficial mutation to increase rapidly upon selection, at the end the whole population becomes fixed for that beneficial mutation (Smith and Haigh.,1974). It results in reduction in genetic diversity in the population with an increase in frequency variants, long homozygous regions, and high levels of linkage disequilibrium (Pritchard et al., 2010).

Soft sweep can be of two types based on origin as single and multiple origin soft sweeps (Hermisson and Pennings., 2017). In the first type, the selection acts on 'single standing genetic variation that previously had either neutral or deleterious effect but has changed and become as beneficial or adaptive due to environment or genetic changes

(Hermisson and Pennings, 2005). This recent advantageous allele trends to increase in frequency towards fixation and leaves some genetic variation. Multiple origin soft sweep often develops in large populations where the occurrence of mutations is quite frequent. These multiple beneficial mutations occur at a single locus on different genomic backgrounds, and they all rise in frequency simultaneously in such a way that none of them can reach fixation level (Pennings and Hermisson, 2006). Here, all the beneficial alleles are usually observed at intermediate frequencies in a population trace back to different mutational origins. The signatures of selection tend to be less pronounced in soft sweeps and it does not drastically reduce the genetic variation in the population (Pritchard et al., 2010). Hence, soft sweeps are more difficult to detect than hard sweeps and the probability of detection depends on the sample size.

2.3.1 Neutral theory

The neutral theory of molecular evolution suggests that the majority of genetic variations are due to neutral substitutions maintained by genetic drift, where all individuals have the same level of fitness to survive (Kimura, 1968). Several recent papers have been using modern statistical approaches to identify differences between neutral and adaptive change in the genome as a prevalence of evolution (Shapiro et al., 2007; Grivet et al.,2017). However, the assumptions and models of genetic drift based on neutral theory served as a backbone for the development of modern statistical approaches used for the detection of selective sweeps (Vitti et al., 2013; Kern and Hahn, 2018). Under the drift model, a common allele should have short haplotypes. But, if a common allele is having a long haplotype, it is considered to have a good signal for selection.

2.3.2. Methods for detection of selection signatures:

2.3.2.1 Evolution of selection scan statistics:

During the 1990's gene substitution-based comparison methods (d_N/ d_S) such as *Hudson Kreitman Aguade (HKA)* test and *McDonald Kreitman test (MKT)* used to find divergence of species. The d_N/d_S (also referred as K_a/K_s or ω) is the ratio of the rate of substitutions at synonymous sites/neutral sites (d_S) to the rate of substitutions at non-synonymous sites/selected sites (d_N) (Goldman and Yang, 1994; McDonald and Kreitman, 1991; Nielsen, 2005).

In case of positive selection, the ratio will be greater than one and in negative selection, the ratio will be less than one. If $d_N/d_S = 1$ the sites will be under neutral

evolution (Goldman and Yang, 1994). These approaches were used to categorize selection incidents that arise in the past, which represents macroevolutionary patterns that occurred as a result of divergent selection between the species (MacEachern et al., 2009). Initially MKT used for detecting footprints representing adaptive evolution between species based on nucleotide divergence. Later it was further developed as an advanced multilocus MKT for comparing within species nucleotide diversity (McDonald and Kreitman, 1991; Egea et al., 2008).

Traditional population genetics was focused on the comparison of a particular set of markers within a region against neutral assumptions, either experimentally or from statistical models. Advances in high-throughput sequencing and SNP genotyping technologies have led ton newer genome-wide scans for the detection of selection signatures among the populations within a species. After the development of Genome-wide SNP array for cattle in 2008 (Matukumalli et al., 2009), there is a clear transition from microsatellite markers to dense SNP data.

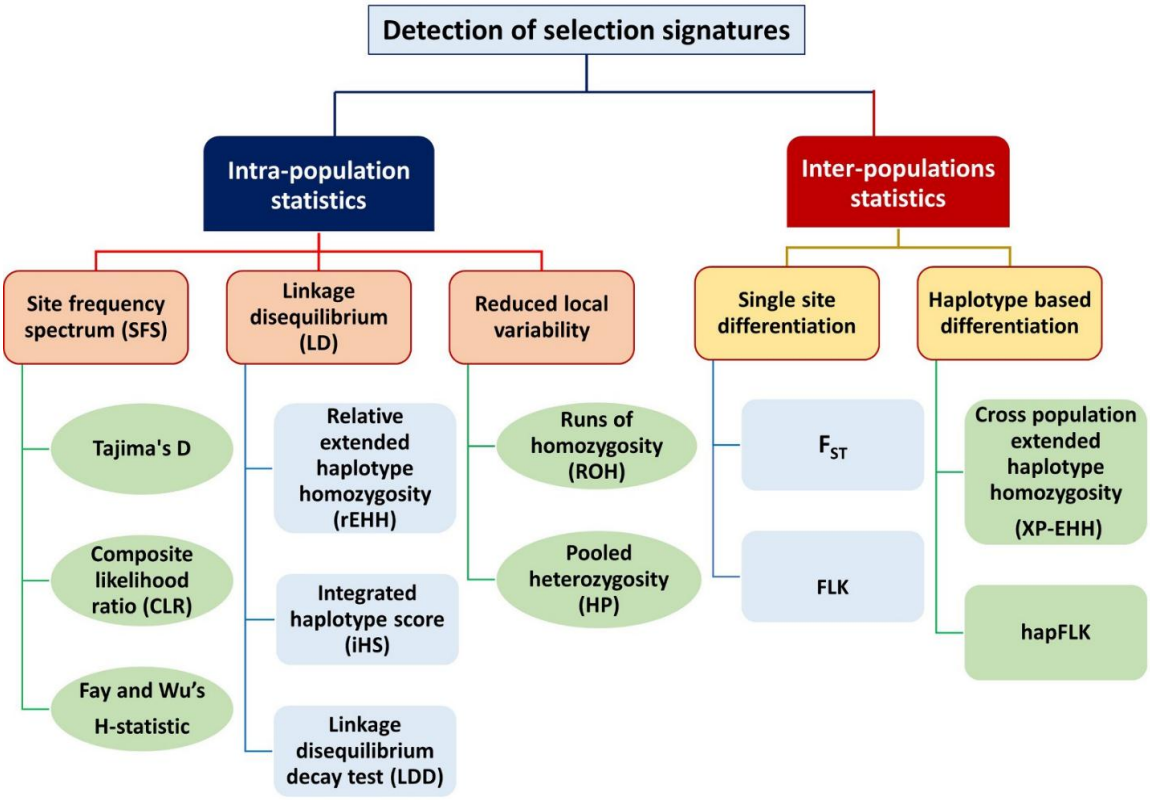
Different statistical methods have been evolved for the detection of selection signatures using SNP genotype data in the livestock populations (Oleksyk et al., 2010; Qanbari and Simianer, 2014). These methods can be broadly classified into two groups: *intra-population statistics* and *inter-populations statistics* (Figure 2.5). Intra-population statistics search for footprints of selection by comparing genomic data within populations. This group includes three primary methods based on the Site Frequency Spectrum (SFS), Linkage Disequilibrium (LD), and reduced local variability (Grossman *et al*., 2010). Inter-populations statistics mainly rely on the degree of differentiation due to locus specific allele frequencies between the populations. These methods can be grouped into Single site and haplotype-based differentiation (Zhao et al.,2015).

2.3.2.2 Intra-population statistics

a) Site frequency spectrum (SFS): Site frequency spectrum (SFS) based on the distribution of allele frequencies in a population. Selective sweeps generate an increased frequency of beneficial variants, and decreased frequency neutral variants (Ronen et al., 2013). Allele frequency-based neutrality tests include *Tajima's D* (Tajima, 1989), *Fay and Wu's H statistic* (Fay and Wu, 2000), and Composite Likelihood Ratio test (CLR) (Lindsay, 1988).

Tajima's D is a classical method that compares the difference between the average number of differences in nucleotides (θ_π) and the number of segregating sites (θ_s) estimated from polymorphism data (Carlson et al., 2005). The *D* value will be zero under neutrality. In the case of positive selection (or selective sweeps), mutations will be rare that reduce heterozygosity and give a negative value of D *(D < 0)*. In contrast, balancing selection leads to intermediate allele frequencies, and the D values will be positive (Tajima, 1989; Simonsen et al., 1995).

Figure 2.5: Different methods of detecting selection signatures



(Source: Saravanan et al.,2020)

Fay and Wu's H statistic is based on the frequency spectrum of ancestral and derived alleles, assuming that ancestral alleles are known. This statistic focuses on the detection of recent positive selection, mainly for non-ancestral (derived) alleles in a medium to high frequency complementary to Tajima's D which can identify low to medium frequency alleles (Cadzow et al., 2014).

Composite Likelihood Ratio (CLR) statistics evaluates the skewness of the frequency spectrum of alleles across the multiple loci and incorporates recombination rate to differentiate selection from demographic events (Nielsen, 2005; Vy and Kim,

2015). The CLR test is based on hypothesis testing that compares the neutral model for the site frequency spectrum of a genomic window with the selective sweep model (Chen et al., 2018). CLR test is extremely sensitive to detect signals of positive selection across multiple sites in a population (Williamson et al., 2007). But these SFS based methods were not suited for locating genome wide SNPs.

b) Linkage disequilibrium-based methods: Selective sweeps creates a haplotype of high frequency alleles with extended Linkage disequilibrium (LD), creates non occurence of recombination during the rapid increase in the frequency of a haplotype carrying a beneficial mutation (Sabeti et al., 2002). LD based methods detects these long homozygous regions with haplotypes of high frequencies generated by selective sweeps. LD based methods are especially useful in the detection of variants under partial or soft selective sweep. Sabeti et al. (2002) proposed method called Extended Haplotype Homozygosity (EHH) based on linkage disequilibrium that detects selection signatures within a population. EHH is the probability that a pair of chromosomes carry homozygous core haplotypes. Then, rEHH (relative extended haplotype homozygosity) is calculated to compare the EHH values of two core haplotypes. Core haplotypes with high rEHH values and high frequency in the population is said to be under positive selection (Sabeti et al., 2007). The relative extended haplotype homozygosity (rEHH) method is used to identify regions that have undergone a recent positive selection and do not require the description of ancestral alleles.

An advance of EHH method, Voight et al., (2006) evolved an integrated haplotype score (iHS) that included recombination distance into calculation. Recent development in genome wide SNP detection technology, have been used to identify of recent selection signatures. iHS is a measure that helps to access how the haplotypes act unusually around an SNP compared to the whole genome (Voight et al., 2006). In this method, each individual SNP is treated as a core SNP and EHH values are calculated for each core SNP according to their ancestral and derived allelic status using a reference genome. Extreme negative iHS values (iHS < -2) indicates more extended haplotypes on derived allelic background compared to haplotypes associated with ancestral alleles, whereas extreme positive values (iHS > 2) suggest that the ancestral allele has swept through the population (Weigand and Leese, 2018).

In comparison with SFS based methods, LD based methods requires haplotype phasing, recombination map, genomic position, ancestral and derived allelic

information for each core SNP. This method can be effectively used to detect signatures when the selected alleles are at intermediate frequencies. Compared to rEHH method, the iHS method is minimally affected by demographic factors and thus the probability of occurrence of false-positive results will be less (Voight et al., 2006).

c) Methods based on reduced local variability: These methods intent to identify genomic regions with reduced variation relative to the genome average such as runs of homozygosity (ROH) (McQuillan et al., 2008) and pooled heterozygosity (H_P) (Rubin et al., 2010).

Runs of homozygosity (ROH) are contiguous lengths of homozygous genotypes that occur within an individual when two haplotypes share a recent common ancestor identical by descent (IBD) (Gibson et al., 2006). ROHs method is used to assess genomic inbreeding levels, population structure, and demography history in livestock populations (Curik et al., 2014). Theoretically, a selective sweep should have stretches of homozygous loci that exhibits greater homozygosity than the average of the genome (Almeida et al., 2019). Hence, ROH can also be used to identify signatures of selection as the individuals that have undergone selective process will exhibit long runs of homozygosity around the target locus (Rebelato et al., 2018). Whereas, pooled heterozygosity (H_P) uses allele counts to calculate heterozygosity and estimates the deviation of expected local heterozygosity depression in chromosomal windows from the average heterozygosity of the genome (Rubin et al., 2010).

2.3.2.3 Inter-populations statistics

a) Single site population differentiation: Single site population differentiation-based methods include Fixation index (F_{ST}) (Wright, 1949) and FLK (Bonhomme et al., 2010). F_{ST} is based on the measure of differences in frequencies of alleles between populations, value ranges from '0' indicates no differentiation between population to '1' indicates a certain amount of fixed difference between populations. Highly differentiated allele frequency between the populations at any given locus (i.e., higher F_{ST} values) indicates positive selection, whereas low F_{ST} values suggest negative selection (Zhao et al., 2015). This method faces a problem of overestimation when sample size is small, but recent SNPs based methods increased the efficiency of detecting genetic differentiation even in case of small sample size (Willing et al., 2012).

Various other F_{ST} estimators' method notably such as-

- Weir and Cockerham's F_{ST} (Weir and Cockerham, 1984)
- Hudson's F_{ST} (Hudson et al., 1992)
- Holsinger's F_{ST} (Holsinger, 2004)
- Population specific F_{ST} (Weir and Hill, 2002)
- Drift model based F_{ST} (Nicholson et al., 2002)
- Bayesian model based F_{ST} (Gianola et al., 2010)

An advantage of F_{ST} over other methods based on LD or SFS is that F_{ST} is SNP-specific and can detect the actual genetic variants under selection. Rather than analyzing each SNP separately, it is more advisable to scan for several consecutive SNPs with average F_{ST} score (by use of genomic windows). When using hierarchically structured data sets, F_{ST} statistics can detect false positive/negative results (Fariello et al., 2014).

FLK (or T_{FLK}) is an extension of Lewontin and Krakauer (LK) statistic (T_{LK}) which compares observed and expected variances of F_{ST} estimated through variance ratio test under neutrality respectively. FLK uses a phylogenetic estimation of the population's kinship (F) where it accounts for changes in effective population sizes over time and incorporates the hierarchical branching of populations (Lewontin and Krakauer.,1973; Bonhomme et al.,2010). FLK is a powerful parametric test that can easily handle massive genotype data sets for the detection of selection signatures among different populations. Compared to the F_{ST} approach, FLK efficiently reduces the type 1 error (false positives) during the detection of selection signatures (Bonhomme et al., 2010).

b) Haplotype based differentiation methods: Haplotype based differentiation methods include cross population extended haplotype homozygosity (XP-EHH) (Sabeti et al., 2007) and hapFLK, a haplotype-based extension of the FLK statistic (Fariello et al.,2013). These methods use haplotype information in multiple populations and SNP ascertainment bias will be less. Cross population extended haplotype homozygosity (XP-EHH) is a haplotype-based differentiation method, introduced by Sabeti et al. (2007). To calculate XP- EHH between two populations, first, iHH values for each population are calculated separately by integrating the EHH of the entire sample in the population. XP-EHH score is directional, and the positive and negative scores suggest that the selection occurred in these population, respectively.

Fariello et al. (2013) proposed a haplotype based FLK method called hapFLK. The hapFLK statistic is based on both haplotype information and hierarchical structure of populations that result in a higher power for the detection of genomic regions under selection. Unlike F_{ST} methods, it accounts for varying effective population sizes. The estimation of the population's kinship (F) matrix is similar to the FLK method, but instead of computing from allele frequencies, the statistics are determined from haplotype frequencies (Fariello et al., 2013). This method can be applied to an unphased SNP genotype data. The hierarchical population structure is estimated by using pairwise Reynolds' genetic distances (Reynolds et al., 1983) between the populations. The pairwise Reynolds' distances between populations are calculated for each SNP and then averaged over the genome and con- verted into the Kinship matrix (Brito et al., 2017).

2.3.3 Application of selection sweeps

Selection sweeps are primary used to identify multiple biological pathway influence production and adaptation. Some studies also aim to find the evidence of positive selection of admixture proportion in crossbred population e.g., Swiss Fleckvieh cattle (Khayatzadeh et al.., 2016), Creole cattle (Gautier and Naves., 2011). It helps in indirect mapping of candidate gene location in selected regions and to understand genetic architecture of livestock species (Druet et al., 2013). Evolutionary pattern of local adaptation of varied species in same geographical region seems hindering the commonality between varied species that are responsible for their unique ability to sustain in these regions (Witt and Sánchez., 2019).

2.3.4 Quantitative Trait Loci

Selected regions were overlapped with reported Quantitative trait loci (QTLs) helps in finding candidate genes in that region and helps to classify genes in trait wise (Gutiérrez-Gil et al., 2015). Recent year greater importance was given to QTLs act as an important genetic marker in selection of livestock. Till date, a total of 1,30,407 QTLs for 634 traits were identified and listed in public domain (www.animalgenome.org). QTLs located in BTA 3, 6, 14, 20 and 26 lodges multiple candidate genes affecting milk production notably ABCG2, B4GALT1, CSN2, CSN3, DGAT1 and LALBA, affects milk yield and milk composition (Khatkar et al.,2004; Ogorevc et al., 2009), similarly BTA 3, 5, 12, 16, 19 and 27 has QTL has candidate genes responsible for respiratory regulation, blood parameter, coat colour etc like EPAS1 for hypoxia , EGLN1 for

angiogenesis, PPARA reduce hypoxia inducing factor, KIT for coat colour etc (Howard et al., 2014).

2.3.5 Selection sweeps studies in livestock:

Selection signatures are extensively studied in recent years and gain importance in animal breeding. Contrast to genetic diversity studies which only analyze the degree of differentiation, whereas selection signature study goes beyond that and detects the actual reason for that diversification among the livestock breeds or populations (Cesarani et al.,2018). Realizing how selection pressure acts on a particular population may aid us to formulate effective breeding plans to improve economically important traits (Gurgul et al., 2018).

a) Cattle: Selection sweeps studies has helped in identifying of less known genes like *LAP3*, *SAR1B*, *LRIG3*, *FGF5* and *NUDCD3* positively selected for milk production and its composition in Chinese Holstein (Xu et al., 2014). Combine selection sweep analysis on 25 breeds has found genes on BTA-6 (*ABCG2*, Casein cluster), BTA-7 (*SAR1B, HBEGF*), BTA-14 (*DGAT1*), BTA-16 (*AGTRAP*, *KIF1B*) and BTA-20 (*GHR*) have been selected strongly for milk production in various dairy breeds (Randhawa et al., 2016). A study on Gir cattle found selection sweep regions has a list of genes associated with production like *KCNIP4, ANXA4, FTO, EGFR, and PAK1* (Maiorano et al., 2018). The candidate genes *MRPS30* and *FGF10* on BTA20 were found to be associated with milk yield, protein percentage, and mastitis resistance (Kadri et al.,2015). Moravčíková et al. (2019) performed selection signature analysis in six beef breeds (Aberdeen Angus, Hereford, Limousin, Charolais, Piedmontese, and Romagnola), found several candidate genes related to double muscling (*MSTN*), muscle development (*GDF9, GHRH, GHR*), intramuscular fat content (*SCD*) and meat tenderness (*CAST*). Signer-Hasler et al. (2017) found more selection signals in the BTA5 which harbour *KITLG*, *STAT6*, and *SYT10* genes in- volved in growth efficiency, carcass traits, and longevity. The candidate genes under selection such as *AQP5* (on BTA5), *RAD50* (on BTA7), and *RETREG1* (on BTA20) were found to be associated with response to cold/heat acclimation in Russian cattle breeds (Yurchenko et al., 2018). Igoshin et al. (2019) found a single candidate selection region on BTA15 harbouring *MSANTD4* and *GRIA4* genes associated with cold-stress resistance phenotype.

b) Sheep and Goat: Purfield et al. (2017) performed a genomic scan for selection signatures in 6 meat purpose sheep breeds. Several genomic regions were identified influencing muscle formation and promote weight gain among those breeds. The *NPR2* gene located on OAR2 is associated with body size and skeletal morphology and has been identified as a candidate gene for selection in some other studies in sheep (Kijas et al., 2012; Manzari et al., 2019). Onzima et al. (2018) evaluated selection signatures six Ugandan goat breeds (Mubende, Kigezi, Small East African, Karamojong, Sebei, and Boer), revealed that several genes within the selected regions were associated with production traits (*GJB2* and *GJA3*), immune response (*IL10RB* and *IL23A*), and thermotolerance (*MTOR* and *MAPK3*). Comparative genomic approaches (XP-EHH and XP-CLR) in ten goat populations using whole genome sequencing data, that uncovered several candidate genes related to Salmonella infection and cardiomyopathy (Kim et al.,2019).

c) Pig: Chen et al. (2018) detected selection signatures in Chinese indigenous Laiwu pigs using Porcine HD SNP data. They identified some of the functionally important candidate genes for feed intake and fat deposition (*NPY1R, PIK3R1,* and *JAKMIP1*), immune response (*CXCL2, CXCL8,* and *TLR2*), and reproduction (*ESR1* and *PTHLH*). 154 putative regions found under selection and QTLs associated with backfat thickness and average daily gain in Duroc pigs (Diao et al., 2018). Similarly, Ma et al. (2019) employed comparative genomic approaches (F_{ST} and XP-EHH) to detect signatures of selection associated with backfat thickness traits in Yorkshire pigs.

d) Horse: Strong selection signals for sports performance in Swedish Warmblood horses found by Ablondi et al. (2019). Nolte et al. (2019) detected many important candidate genes under selection associated with muscle functionality, energy metabolism, growth, and fertility in four German Warmblood horse breeds.

e) Chicken: *TSHR* gene selected in domestic chickens which has a vital role in metabolic regulation and photoperiod control (Rubin et al., 2010). *RB1*, *BBS7*, *MAOA*, *MAOB, EHBP1*, *LRP2BP*, *LRP1B, MYO7A*, *MYO9A* and *PRPSAP1* were detected positively selected in broiler which are involved in abdominal fat deposition (Zhang et al.,2012).

f) Yaks: *GRIK4, IFNLR1 and LOC102275985* genes found selected in tibetan yaks influencing glutamatergic synapse, JAK-STAT signaling pathway and olfactory

transduction pathways involved in environmental adaptability and physiological regulation in hypoxic conditions (Guang-Xin et al., 2020). Jinchuan yak possessed 339 significantly and positively selected genes involved in physiological rhythm and the breed's excellent production characteristics (Lan et al., 2018).

2.3.6 Selection signatures related to high altitude conditions:

Tremendous selective pressure imposed by the high-altitude environment caused convergent adaptations at both the molecular and phenotypic levels that have developed across species (Graham and McCracken 2019; Natarajan et al., 2015). Few genetic variants underwent convergent evolution across all species, according to a comprehensive study comparing populations from the Tibetan plateau between humans and livestocks. Wu et al., (2018) proposes that rather than single genes, evolution works on genes implicated in comparable functional pathways within a network.

According to Witt and Huerta-Sànchez (2019), observable physiological changes in altitude adaptation are similar in humans and livestock animals, but the genetic basis for the adaptation differs between populations. In cattle, selection pressure for high altitude differs by geographical region: in Ethiopia (Edea et al., 2014), BDNF, TFRC, and PML genes were identified as top candidates for the hypoxia response, whereas in Tibetan cattle, EGLN1 and HIF3A genes were identified as top selection targets for the hypoxia response (Wu et al., 2018). However, several common genes with positive selection signals across species, such as EPAS1, JAZF1, DKK2, and SPON1, have adaption fingerprints in Tibetan plateau cattle, sheep, and goat populations, implying convergent evolution at the molecular level (Wu et al., 2020). EPAS1 and HIF2A are the most well-known hypoxia-related genes with adaptive characteristics in a variety of species of high altitude. These genes are more likely to be targeted by selection as a major transcription factor of the HIF pathway (Martin and Orgogozo ., 2013).

2.3.7 Bioinformatics tools for identifying selection signatures

The general workflow for the detection of selection signatures starts from the generation of genotype data, either using SNP microarray or high-throughput sequencing technologies. This raw genotype data is processed according to the method of analysis and bioinformatics tools used, as each of them requires specific input formats (Cadzow et al., 2014). Many selection sweep detection software are available in open source, majority of the programs work in the Linux platform, some are suitable for use in

Windows, but few have the version for macOS. The commonly applied bioinformatics tools for the detection of selection signatures in livestock populations are listed in Table 2.5.

Table 2.5: **Bioinformatics tools for use to detect selection signature**

Program/ tool	Latest version	Methods	Source link
Arlequin	3.5.2.2	Tajima's D	*http://cmpg.unibe.ch/software/arlequin35/*
BayeScan	2.1	Fst	*http://cmpg.unibe.ch/software/BayeScan/index.html*
DetectRUNS	0.9.6	ROH	*https://CRAN.R-project.org/package=detectRUNS*
Hapbin	1.0	EHH, iHS, XP-EHH	*https://github.com/evotools/hapbin*
HapFLK	1.4	HapFLK,FLK	*https://forge-dga.jouy.inra.fr/projects/hapflk/files*
Omega Plus	2.2.2	LD based	*http://www.exelixis-lab.org/software.html*
PLINK	1.90	ROH,Fst	*http://zzz.bwh.harvard.edu/plink/*
PopGenome	2.7.5	Tajima's D, Fst (Bayesian approach)	*https://CRAN.R-project.org/package=PopGenome*
Rehh	3.1.0	EHH, iHS, Rsb, XP-EHH	*https://CRAN.R-project.org/package=rehh*
Selscan	1.2.0	EHH, iHS, XP-EHH, nSL	*https://github.com/szpiech/selscan*
SweeD	3.2.1	CLR	*https://cme.h-its.org/exelixis/web/software/sweed/index.html*
Sweepfinder2	1.0	CLR	*http://degiorgiogroup.fau.edu/sf2.html*
VariScan	2.0.3	Tajima's D, Fu, and Li's tests	*http://www.ub.edu/softevol/variscan*
VCFtools	1.16	Tajima's D, Fst	*https://github.com/vcftools/vcftools*
XP-CLR	1.1.2	XP-CLR	*https://reich.hms.harvard.edu/software*

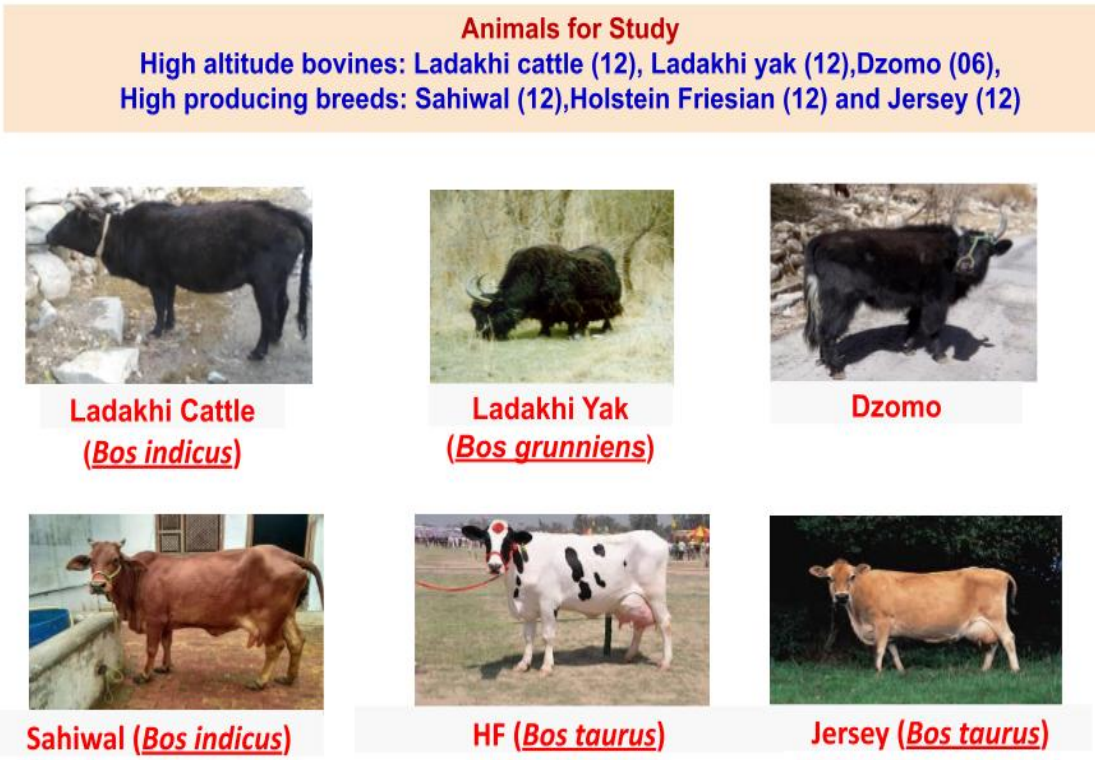
CHAPTER –3

Materials & Methods

3.1 Study population:

Primary study group comprises high altitude bovines namely Ladakhi cattle (n=12), Ladakhi Yak (n=12) and Yak X Cattle hybrid (n=6) of Ladakh region one of the highest regions in the world. It is in the upper Himalayas at altitudes ranging from 3000 to 5000 m above mean sea level (msl) and the hardest climatic zone of the Indian Sub-continent. Secondly the well-known Indian and Exotic milch breeds notably Sahiwal (n=12), Jersey (n=12) and Holstein Friesian (n=12) were used as comparison study populations with high altitude bovine group (Figure 3.1).

Figure 3.1: Pictures of high altitude and high producing bovine groups under study



3.2 Blood collection

Blood samples of 10 ml volume were collected aseptically by jugular vein puncture from respective study groups using sterile 0.5 % EDTA vacutainer tubes for the isolation of the genomic DNA and labelled samples were stored in deep freezer (-20^{0}C) until further processing.

3.2.1 Genomic DNA extraction

Genomic DNA was isolated by Phenol-Chloroform-Isoamyl alcohol (PCI Method) (Russell and Sambrook, 2001).

a) RBC lysis

- Initially freshly prepared 10ml of RBC lysis buffer was added to 5 ml of sample blood in a centrifuge tube and kept in chilling conditions for 5 minutes. Followed by centrifugation at 10000 rpm for 10 minutes and supernatant was discarded.
- Again, 5ml of RBC lysis buffer was added and centrifuged again for 10 minutes. Same step it repeated till a white-colour pellet was observed at bottom of tube,
- Vortexing was done to mix the contents well followed by centrifugation at 10,000 rpm for 10 minutes and supernatant were discarded

b) DNA Extraction

- 5 ml of digestion buffer ,125 µl of 20% SDS and 25 µl of Proteinase-K were added to tubes and kept overnight at 56ºC.
- 5 ml of tris-equilibrated phenol was added and left for 5 minutes as in room temperature.
- Later, these tubes were centrifuged at 8000 rpm for 10 minutes. Upper floating phase content was transferred to another tube. Chloroform and Isoamyl alcohol were added to these phases in ratio of 24:1 and left to react for 5 minutes. Later, they were centrifuged at 8000 rpm for 10 minutes, at room temperature.
- Aqueous phase content was transferred to a new Falcon tube. Solution 100 µl of Sodium acetate, followed by 200 µl of chilled Ethanol was added per ml of aqueous phase, respectively. Then these tubes were kept at - 4^0C to precipitate DNA.
- The spooled threads of g-DNA were transferred to Eppendorf's tubes. 1 mL of 70% Ethanol was added to those tubes and centrifuged at 7000 rpm for 5 minutes at 4 ^{0} C. Discard and repeat ethanol step one more time till g-DNA settle as pellet.
- Finally add 250 µl of TEB was added at 37^0C to dissolve DNA pellet. Dissolved DNA samples were stored in deep freezer.

c) Quality of DNA samples

Agarose gel electrophoresis was carried out, with 0.8% Agarose solution, with a potential difference of 50 Volts applied across the electrodes for an hour. Gel was seen using UV illuminator to check for quality of bands and smearing.

d) Quantity of DNA samples:

It was evaluated using nanodrop spectrophotometer. Optical density (OD) at 260 and 280 nm wavelengths were recorded

Concentration of DNA (µg/ml) = OD 260 X 50µg/ml X Dilution Factor

Yield of DNA (µg) = DNA concentration X Total sample volume

The samples passing quality and quantity check were further processed.

3.3 Construction of DNA libraries

Library construction was done as per the protocol mentioned in Peterson *et al.* (2012).

- 1 µg of each DNA sample was subjected to double digestion by the two set enzymes (SphI, EcoRI for 5') and (MluCI, MseI for 3') in a digestion buffer at 37º C for 3 hours.
- With the help of AMPure XP beads, the double-digest was cleaned and checked for quantity. P1 adapter (carrying a unique barcode) and P2 adapter (carrying a common barcode) were added using T4 DNA ligase, with a master mix.
- The ligated products were then pooled up for size-selection and later cleaned using the beads.
- The digesta was then run on 2% Agarose Gel Electrophoresis (AGEP) and further, size selection was carried out.
- Illumine adapters and flow cell annealing sequences were incorporated into the size selected DNA sequences, using PCR to increase overall concentration of libraries.
- Quality Control was done using Bioanalyzer.
- Libraries were pooled by combining them in equimolar concentrations for sequencing.

3.4 Sequencing and raw read generation:

Samples will be loaded into flow cells after library preparation and subjected to sequencing using illumina Hiseq 2000 by means of synthesis sequencing, which will generate short read sequences with unique product size up to 151 bp length.

3.5 Bioinformatics analysis

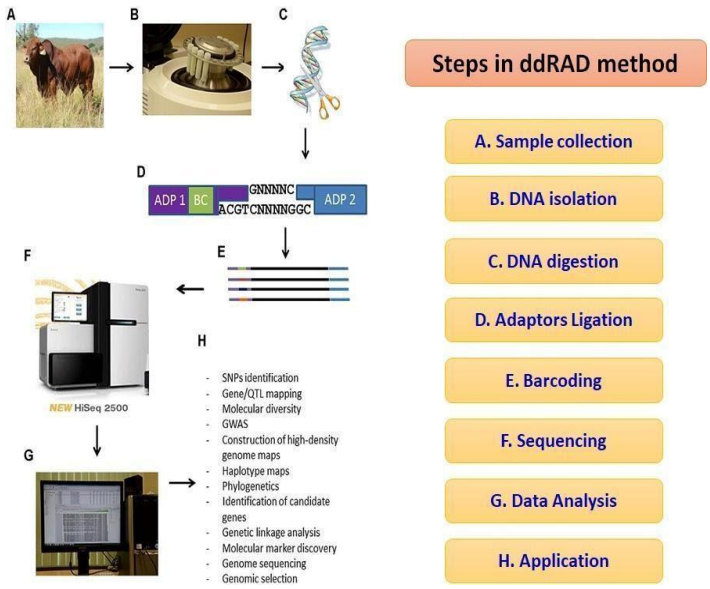
3.5.1 Demultiplexing the reads

Sequencing data obtained from the sequencing firm was demultiplexed using FastX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Each of the demultiplexed reads was 151 bp long.

3.5.2. Quality check:

Raw reads will be accustomed to preliminary processing such as trimming raw reads where the sequence adapters will be filtered out. Quality control will be carried out determine read quality, GC content proportion and size of read. Preliminary screening of raw reads was done using *FastQC* (Andrews., 2010). *Prinseq lite* 2.0 version (Schmieder and Edwards., 2011) was used to generate graph data and graphs of the sequenced raw reads. Going through the graphs, corresponding restriction ends were located and all the barcode and adapter sequences falling before recognition sites were removed. *STACKS* 2.3 (Catchen et al., 2013) *process rad tags* feature was used to locate RAD loci and to remove reads having phred score below 15.

Figure 3.2: Schematic diagram of raw data generation using ddRAD sequencing



3.5.3 lignment against reference genomes

Reference genome's FASTA, GFF and GTF files were retrieved from NCBI database (https://www.ncbi.nlm.nih.gov/) for mapping with sample sequence. Bowtie2 (Langmead et al., 2009) was used to create reference genome index and map the query sequences under investigation with reference index, in *local very sensitive* alignment mode with a minimum mapping quality of 30 to create Sequence Alignment Map (SAM) file.

3.5.4 Post-Alignment processing

SAM files were further processed to Binary Alignment Map (BAM) files using SAM tools (Li et al., 2011). BAM files were further sorted, index and finally mpileup into a single Binary variant Call Format (BCF) file using BCFtools (Li., 2011).

3.5.5 Variant detection

VCF files were generated from their binaries, filtered at minimum read depths of 2, 5 and 10 and a mapping quality of 30, in order to minimize false positives and later SNPs and INDELs were called using VCFtools (Danecek et al., 2011).

3.5.6 Quality control of variants

Quality check was performed on SNPs obtained at a minimum read depth of 10 for Hardy Weinberg equilibrium (0.001), Minor allele frequency (0.01), Missing genotypes (1.0) and linkage disequilibrium (0.5) in order to obtain High quality SNPs using *VCFtools (*Danecek et al., 2011).

3.5.7 Annotation

SNPs were annotated using SnpEFF in the pipeline (Cingolani et al.,2012). It annotates and predicts the effects of genetic variants on genes (such as amino acid changes), gives detailed reports on distribution of variants in genomic regions and gives amino acid substitution rate.

3.5.8 Detection of individual Selective sweeps:

Detection of selective sweeps was done by a combinatorial approach involving calculation of Site Frequency Spectrum (SFS) and iHS methods, common genes from both methods were used for further downstream analysis. *SweeD 3.0 (Sweep Detector)* is an open-source tool for the Site Frequency Spectrum (SFS) based rapid detection of selective sweeps at whole genome level (Pavlidis et al., 2013).

Integrated Haplotype Score (iHS) by Voight et al., (2006) works based on haplotype homozygosity between ancestral and derived allele. *rehh* a *R* package (Gautier and Vitalis., 2012) was used to analysis iHS score. Samples analyzed breed-wise to detect unique selection sweeps. Top 1 percentile of selected sweeps based on Composite Likelihood and iHS scores were further annotated using UCSC table (https://genome.ucsc.edu/cgi-bin/hgTables) and NCBI gene browsers (https://www.ncbi.nlm.nih.gov/genome/gdv/). Common genes from both methods were sorted used for downstream analysis

3.5.9 Genome wide Co-selected signature detection:

Selection signature between two breeds were analyzed using by combination of cross-population Extended Haplotype Homozygosity (XP-EHH) (Sabeti et al., 2007) and RsB (Tang et al., 2007) former works on Haplotype homozygosity and later works on allele frequency spectrum. Common selected sweeps by both methods are sorted and used to create a co-selected regions with 100 kb window size consisting of a minimum of 2 selected markers. Significant selected regions ($p<0.05$) were further used for QTLs analysis and gene annotation.

3.6 Gene ontology analysis:

Selected gene were used further subjected to Gene ontology using Pather Go online portal (Thomas et al., 2003) where its biological process, molecular function and cellular components were found and used for further analysis. Vital pathways involved by selected genes were found using KEGG pathway (Kanehisa and Goto., 2000). Phenotype ontology analysis was carried out using go:Profiler (Reimand et al., 2007) to describe the phenotype change occurred in breeds due to missense mutation.

3.7 Gene to Gene interaction network:

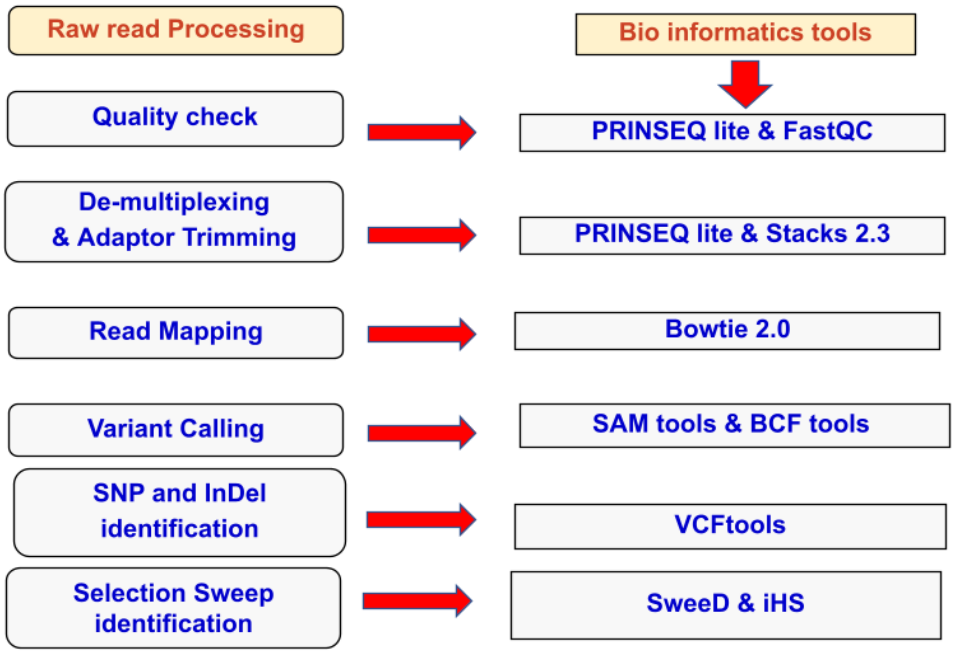
Selected genes were used to construct an interactive gene network based on its adaptive biological regulation to hypoxic condition, cardiovascular system, cutaneous and energy metabolism using ClueGo (Bindea et al., 2009) of cytoscape platform.

3.8 Identifying the Co-selected regions and QTL association:

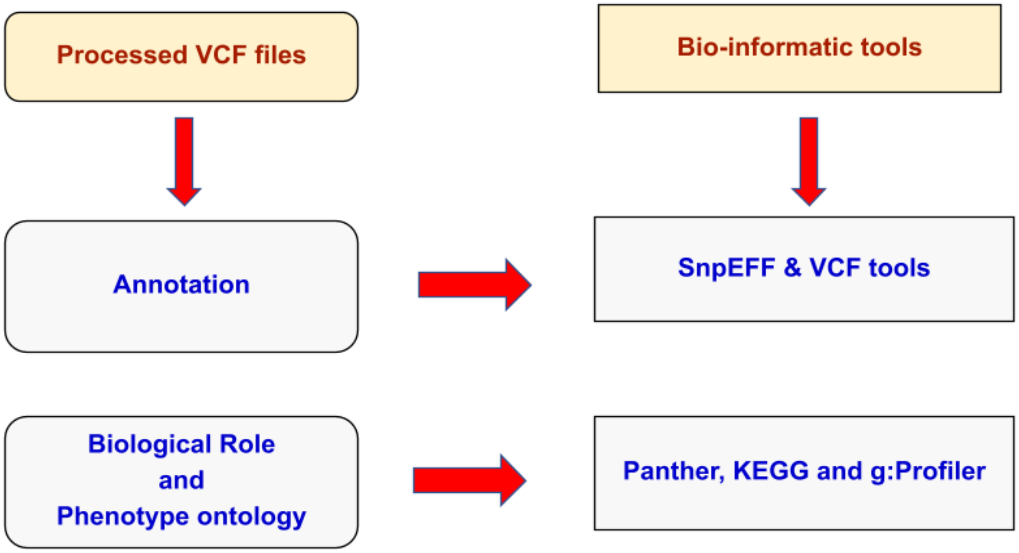
Co-selected regions among compared groups were QTL mapped using Cattle QTL database (Hu and Reecy., 2007). Further those are grouped as Milk related (Milk yield, Milk Fat yield, Milk Protein Yield, Milk fat percentage, Milk Protein percentage and Milk components), Reproduction related (Fertility, Conception rate, Calving ease, Spermatozoa morphology and fertility index) and Adaptation related (Skeletal morphology, Physical Stature, Blood morphology, Any special adaptation such as cold tolerance, Skin coat and Immunity). The comparative groups with optimum numbers of co-selected regions among these three traits may be assumed to have complementary nature between those two breeds.

Figure 3.3: **Objective-wise flow chart of Bioinformatic pipeline**

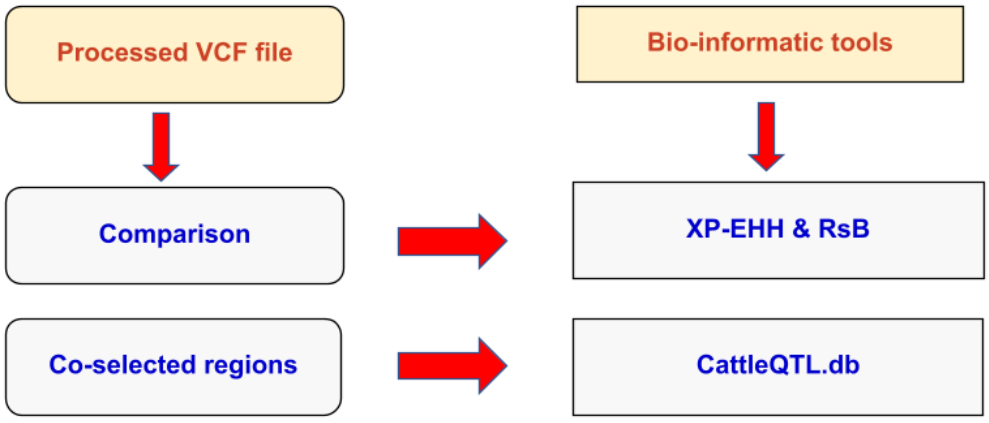
Objective 1: To identify genetic variants and selection sweeps in Ladakhi cattle, Yak, and their hybrids



Objective 2: To functionally annotate genetic variants and selection sweeps for milk production and adaptation in high altitude bovines



Objective 3: To compare identified selection sweeps of high altitude adapted bovines with high producing *Bos taurus* and *Bos indicus* cattle



CHAPTER –4

Results and Discussion

RESULT AND DISCUSSION

Ladakh (UT) is the home of many unique and locally adapted dairy animal species like Ladakhi cattle, yak, and their cross hybrids, which are significantly contributing in the form of milk and milk products to the native people in this high-altitude region. The unique genepool of these native bovines is an interesting venture, not only to study the evolutionary footprints of adaptation; but also, important to explore the genomic architect for milk production and adaptation traits. Vis-à-vis, the genomic comparison of these dairy animals with high-producing zebu and taurine cattle breeds may be important to understand the complementarity for crossbreeding, without affecting the genetic architect for adaptation. This study tried to find novel genome wide positive selection signatures in high altitude bovines (Ladakhi cattle, Ladakhi yak and yak X cattle hybrid) through which the genetic footprints responsible for adaptation in terms structural and physiological to the cold hypoxic condition of high altitude could be explored. Secondly, co-selected signatures among high altitude bovines (Ladakhi cattle and yak) with high producing cattle (Sahiwal, Holstein Friesian and Jersey) were found and the possible genetic relationship between two groups was detected by pairwise comparison. From the co-selected signatures, shared selected regions (100kb window) were constructed which are subjected for QTL analysis to sketch out the potent economic traits and vital adaptive traits between high altitude bovines and high producing cattle breeds. Trait annotated shared regions were enumerated and compared pairwise to find proximal genetically compatible breeds that could be used for breed improvement in high altitude conditions. The study indicated the common and converging genomic adaptation among high altitude bovines, probably due to the stringent selection process undergone to cope up with harsh conditions of high altitude. Although hybridization was seemingly found to affect the genomic composition, however, hybrids had more proximity with yak particularly for the regions associated with high altitude adaptation. Further, comparative analysis of high-altitude bovines with high producing cattle indicated a larger number of commonly selected regions between Ladakhi cattle and Sahiwal, providing the greater possibility for better genomic compatibility. The results under the study are described in this section as follows-

4.1 Pre-processed genomic raw read analysis and mapping:

A total of 28.3 and 49.8 Gb of raw data, comprising 80.46 and 108.55 million of reads were sequenced from high altitude bovines (Ladakhi cattle, Ladakhi yak and Yak hybrid) and high milk producing cattle (Sahiwal, Holstein Friesian and Jersey) groups, respectively. After quality check and trimming of the raw data, a total of 78.49 and 105.5 million clean reads were obtained from both of the groups. Overall alignment rate was about 91% for high altitude bovines and 97 % for the Sahiwal, after comparison with *Bos indicus* genome. The Holstein Friesian and Jersey showed 99 % alignment rate with *Bos taurus* genome (Table 4.1). The reference genome of the *Bos indicus* was used in high altitude bovines due to non-availability of chromosome-wise yak reference genome, which might be the reason for lower alignment rate of Ladakhi cattle, yak and their hybrids.

Table 4.1: **Details of sequence reads and alignment rate of high-altitude bovine and cattle breeds**

Bovine species/ Breed	**No. of samples**	**Reference Genome**	**Total data obtained (Gb)**	**Total raw reads (million bp)**	**QC passed reads (million bp) (%)**	**Overall alignment rate (%)**
Ladakhi cattle	12	*Bos indicus*	9.6	39.24	38.23 (97.42 %)	95.11
Ladakhi yak	12	*Bos indicus*	7.8	29.46	28.63 (97.18 %)	86.17
Yak X Cattle Hybrids	6	*Bos indicus*	5.4	11.78	11.23 (95.33%)	90.32
Sahiwal	12	*Bos indicus*	8.7	36.78	34.08 (92.64%)	97.23
Holstein Friesian	12	*Bos taurus*	11.2	42.92	41.51 (96.72%)	99.00
Jersey	12	*Bos taurus*	8.5	28.85	26.66 (92.94%)	99.00

4.2 Identification of genetic variants in high-altitude bovines

The total number of genetic variants at different read depths (RD) 2, 5 and 10 were identified in high altitude bovines (Table 4.2). The percentage of novels SNPs were 59.89% (3,58,286), 86.93% (3,11,045) and 81.26% (2,52,582) at RD 10 in the Ladakhi cattle, Ladakhi yak and Yak hybrid, respectively. The High-quality SNPs at RD 10 were further filtered using minor allele frequency (< 0.05), Hardy-Weinberg equilibrium (p< 0.0001) and maximum (100%) missing genotype, obtaining 498452 high quality SNPs in the Ladakhi cattle, 218470 in the Ladakhi yak and 267958 in the yak X cattle hybrid. The genetic variants in Ladakhi yak were found to be different from earlier reported by Sivalingam and coworkers (2020); in light of the use of different reference genomes. The Ladakhi cattle was found to have a greater proportion of novel SNPs among the bovines, which indicates its uniqueness and adaptation towards harsh climatic conditions.

Table 4.2: **Genetic variants in high altitude bovines at various read depth (RD)**

Group	Variants	RD 2	RD 5	RD 10
Ladakhi cattle	SNPs	767176	711755	610058
	InDels	87960	79694	66617
	Total	855136	791499	676675
Ladakhi yak	SNPs	445993	403835	357981
	InDels	41722	37334	32914
	Total	487715	441169	390895
Yak cattle hybrid	SNPs	404324	368901	311100
	InDels	37500	33632	27747
	Total	441824	402533	338847

4.3 Variant distribution in high altitude bovines:

Identified genetic variants after annotation showed that the majority of variants (50.09%) was present in the intron region and the rest are in intergenic (39.25%), upstream (4.36%) and downstream (4.36%) regions respectively Figure 4.1. About 1% variants were found in exonic regions, of which the majority were synonymous followed by

missense annotated variants 0.76% accounts for coding sequence variants (12,711) majorly comprising 53.4% synonymous variants and 34.8% missense variants (Figure 4.2). The transition (71.92 %) and transversion (28.18%) substitution ratio for annotated SNPs is 2.56 and their predominant genotypes are GA and CT. Identified SNPs transition per transversion ratio was 2.56, where G/A and C/T substitutions were found higher than AT genotypes which resemble findings by workers using a similar kind of reduced representation sequencing (Surya et al., 2019; Kumar et al., 2020; Sivalingam et al., 2020).

Figure 4.1: Genome wide genetic variants distribution in high altitude bovines

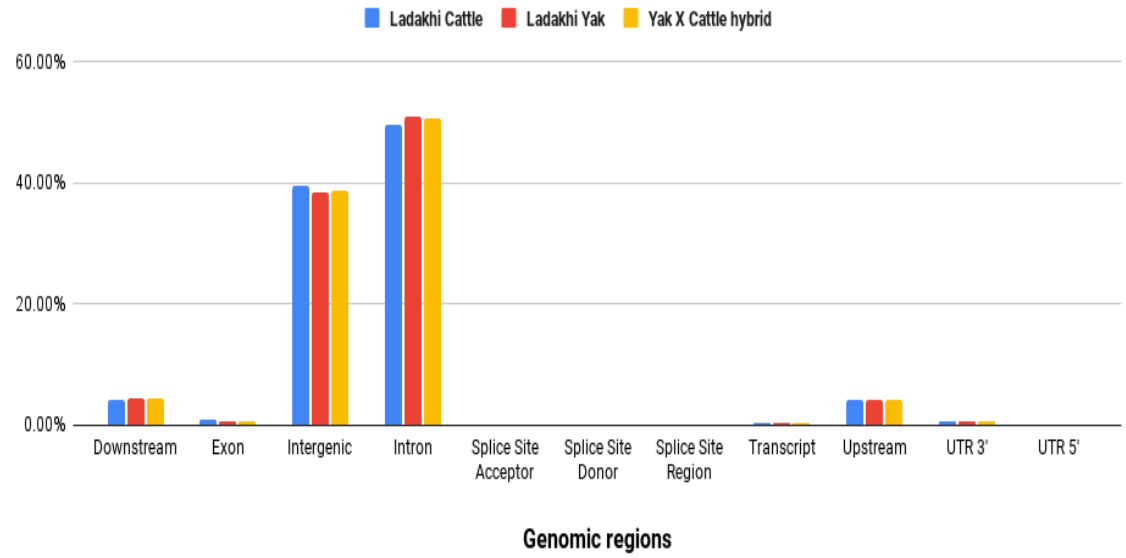
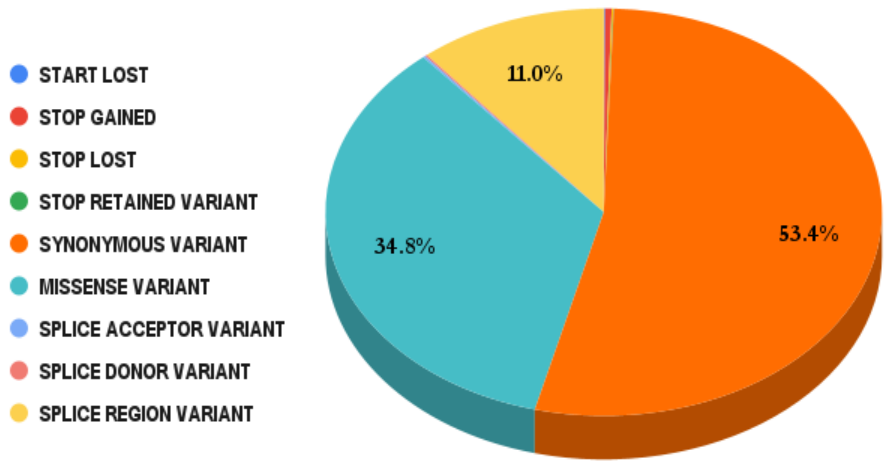


Figure 4.2: Exonic variants distribution in high altitude bovines



4.3.1 Annotation of missense variants:

A total of 2848, 1087 and 857 missense variants were further mapped on 894, 410 and 362 genes in the Ladakhi cattle, Ladakhi yak and Yak cattle hybrid, respectively. Furthermore, phenotype ontology analysis revealed that these missense variants were involved in alteration of structural, physiological, nervous, internal organ's structure and mechanism and cutaneous systems. These missense variants might be helping the bovines to adapt to the high-altitude conditions (Table 4.3).

Table 4.3: Phenotype ontology and no. of annotated genes in high-altitude bovines

Phenotype ontology	No. of annotated genes			P-Value
	Ladakhi cattle	Ladakhi yak	Yak X cattle hybrid	
Alteration of the nervous system	135	98	75	0.0001
Alteration of the musculoskeletal system	128	91	73	0.0001
Alteration nervous system physiology	125	92	71	0.0001
Alteration of the skeletal system	113	90	58	0.0001
Alteration of the digestive system	103	69	52	0.0001
Alteration of the cardiovascular system	96	65	53	0.0001
Alteration of metabolism	81	64	53	0.0001
Alteration of the skin	72	45	40	0.0001
Alteration of the respiratory system	71	43	37	0.0001
Alteration of the genitourinary system	56	43	37	0.0001

Among these, a total of 56 genes were found to be common in high altitude bovines. Notably, polycystic kidney and hepatic disease 1 (PKHD1), BRCA1 interacting helicase 1 (BRIP1), galactosidase, beta 1 (GLB1), adenomatous polyposis coli (APC), fraser extracellular matrix complex subunit 1 (FRAS1), leucine rich repeat containing 6 (LRRC6), methylenetetrahydrofolate dehydrogenase cyclohydrolase and formyltetrahydrofolate synthetase 1 (MTHFD1), NLR family pyrin domain containing 1 (NLRP1), radial spoke head component 4A (RSPH4A) and Tet methylcytosine dioxygenase 2 (TET2) genes were involved in alteration of skeletal and skin coat texture. Spectrin Repeat Containing Nuclear Envelope Protein 2 (SYNE2), collagen type IV alpha 3 chain (COL4A3) and WD repeat containing planar cell polarity effector (WDPCP) gene were involved in alteration of respiratory and cardiovascular system in high altitude bovines.

4.4 SNP density analysis in high altitude bovines:

Among high altitude bovines, the Ladakhi cattle possessed the highest SNP density, with one SNP at every 4366 nucleotides, followed by the Ladakhi yak (one SNP at every 7488 nucleotides) and the yak cattle hybrid (one SNP at every 8563 nucleotides). The chromosomes of high-altitude bovines with high SNP density are listed in Table 4.4.

The SNPs dense regions were calculated at 100 kb non overlapping windows, 253, 293 and 300 regions were found to have high SNP dense patterns in the Ladakhi cattle, yak and yak cattle hybrid, respectively. Among these, a total of 78 regions were common between the Ladakhi cattle and the Ladakhi yak, 85 regions between the Ladakhi cattle and the yak-cattle hybrid and 151 regions between the Ladakhi yak and yak cattle hybrid. The shared proportion of dense regions implies a similar pattern of the SNP density in the Ladakhi yak and the yak hybrid than that of the Ladakhi cattle and the yak hybrid, which reflects more genomic closeness between the yak and the hybrids. Chromosome-wise SNP dense pattern in high altitude bovines is given in *Figure 4.3*. A total of 59 SNP rich regions were found common in these three high altitude bovines (Table 4.5).

Table 4.4: **Chromosomes with high SNP density (top 5) in high altitude bovines**

Ladakhi cattle		Ladakhi yak		Yak X cattle hybrid	
Chromosome	**No. of SNPs**	**Chromosome**	**No. of SNPs**	**Chromosome**	**No. of SNPs**
Ch 24	3366	Ch 25	6234	Ch 25	7239
Ch 23	3579	Ch 28	6513	Ch 12	7512
Ch 12	3590	Ch 27	6586	Ch 28	7589
Ch 28	3743	Ch 20	6658	Ch 22	7642
Ch 27	3844	Ch 24	6658	Ch 20	7690

4.4.1 QTL mapping on SNP dense regions in high altitude bovines

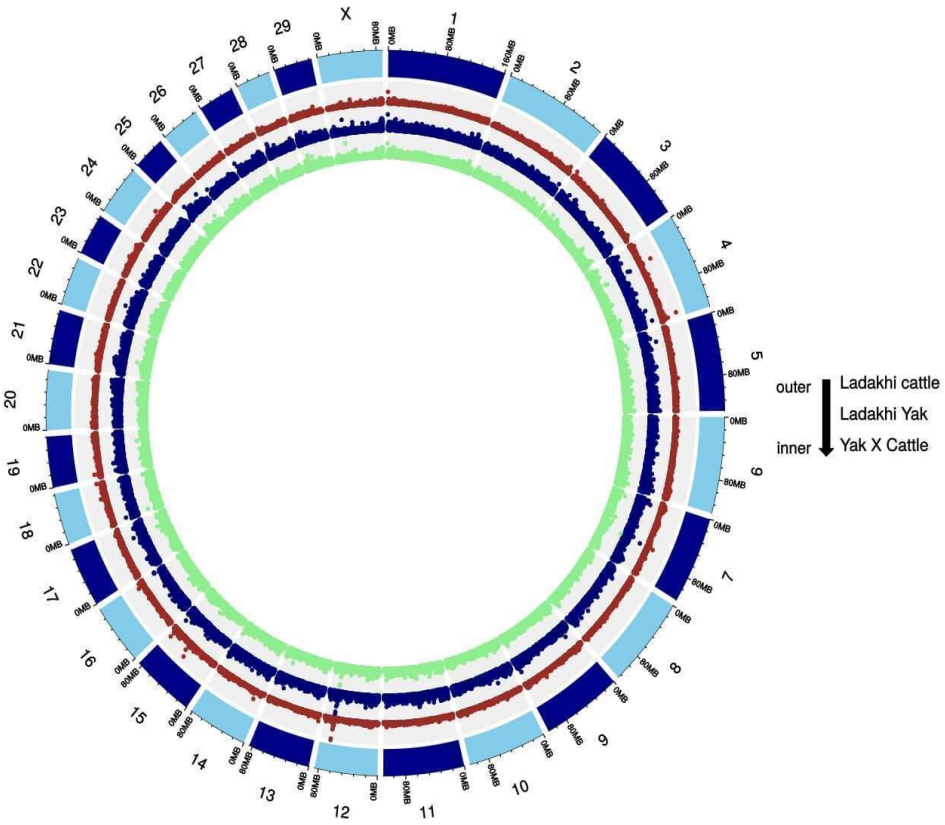
The 59 common high SNP dense regions identified in high altitude bovines were further mapped for presence of *QTLs* associated with the milk (39), reproduction (37) and adaptation (24) Figure 4.4 and Table 4.5. Some of top notables are 70.0 - 70.39 Mb region in the chromosome 12 associated with milk composition, 68.4 - 68.49 Mb region in chr:7 related to cold tolerance QTL, 1 - 1.0 Mb region in the chromosome 1 affects conception rate and 43.9 - 43.99 Mb region in the chromosome 17 affects milk composition and tick resistance. A total of 16 high dense regions comprising all three traits are described in Table 4.6.

Higher polymorphism in these specific regions suggests long-term response to survival and reproduction in the harsh environment. In this study some vital traits such as structural soundness, haematological parameters, calving ease, cold tolerance, and fertility aid to survive and thrive in high altitude conditions were located in regions of high SNP density regions. Hence, these findings suggest that there is continuous genome plasticity occurring in high altitude bovines.

The studies in humans showed difficulty in giving birth and reduced birth weight of young ones at higher altitudes associated with the same genomic/QTL regions (Moore., 2003; Zamudio., 2007; Tal., 2012; Roland et al., 2014). In our study, the highly polymorphic regions in chromosomes- 4, 7, 10, 11, 16, 24, 25 and 26 (Table 4.5) were

related to QTLs conferring calving ease and birth weight. This might help to unfold an evolutionary response in bovines to overcome the parturition difficulties and neonatal growth.

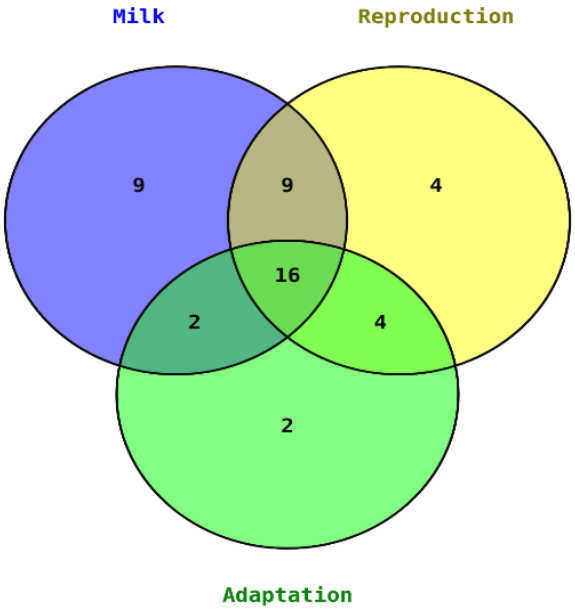
Figure 4.3: Chromosome wise SNP density map in high altitude bovines



(Chromosome wise SNP density map represented from outer to inner by Ladakhi cattle (brown), Ladakhi yak (blue) and Yak cattle hybrid (olive green). Spiked dots are considered highly dense SNP regions)

Cold tolerance is another vital component to survive at high altitude through maintaining normal energy production, optimizing nutritional assimilation, preventing heat loss and also enhancing heat production (Han et al., 2002; Wang et al., 2011; Manou-Sthathopoulou et al., 2015). In this study, 3 high dense SNPs regions in chromosome 7 (Table 4.5 and 4.6) to possess were found common in all three bovine populations of high altitude, indicating that these unique traits might have evolved over ongoing adverse climatic changes to counter the adverse cold stress

Figure 4.4: QTL mapping of common high SNP dense regions in high altitude bovines



Structural soundness is a vital component to survival in the high-altitude conditions, notably the skeletal, foot and leg conformation aids in normal locomotion in steep mountain range for scavenging of food and traits related to stature, skin coat and texture helps to withstand cold stress and higher UV radiation (Storz et al., 2010; Jeong et al., 2018; Fredrich and Wiener., 2020). In this study, 17 high dense SNP regions from various chromosomes (3, 4, 6, 7, 10, 11, 14, 17, 20, 25, 26, 2 and 29) (Table 4.5) were found common in all three high altitude bovines, this higher number of structural soundness related SNP dense regions suggests that there is continuous gene plasticity occurring in high altitude bovine genome to have better adaptation towards high altitude conditions.

Table 4.5: Common high dense SNP regions and corresponding SNP count with their associated traits in high altitude bovines

S.N.	Chr.	Start	End	No. of SNPs			Traits			
				Ladakhi cattle	Ladakhi yak	Yak Hybrid	Milk related	Reproduction	Physiological	Physical fitness
1	1	0	99999	214	75	66		Conception rate		
2	1	9100000	9199999	88	52	41	Milk protein yield, Milk yield, Milk fat yield, Milk protein percentage, Milk fat percentage,	Daughter pregnancy rate		
3	2	3900000	3999999	71	64	43	Milk fat yield,	Age at puberty, Maturity rate, Interval to first estrus after calving	Blood creatinine level, Conjugated linoleic acid content, Lung percentage	Thurl width,
5	3	121400000	121499999	89	48	40		Daughter pregnancy rate		
4	3	28500000	28599999	110	46	47	Milk fat percentage, Milk protein yield, Milk protein Percentage, Milk fat yield			Height

6	3	121700000	121799999	74	46	40				
7	4	7900000	7999999	85	42	40				
8	4	15800000	15899999	90	49	49		Scrotal circumference	Residual feed intake	Stature, Height
9	5	73000000	73099999	91	47	46	Milk yield and composition	Ovulation rate, Scrotal circumference, Heifer pregnancy		Chest depth, Rump length
10	6	59000000	59099999	75	51	43	Milk yield, Milk fat yield, Milk composition	Scrotal circumference, Calving ease, Female Fertility Interval to first estrus after calving, Age at puberty		Stature, Structural soundness
11	6	121700000	121799999	74	46	44				
12	6	117200000	117299999	83	44	41				
13	7	68400000	68499999	139	89	108			Cold tolerance	Height
14	7	7600000	7699999	100	71	65	Milk protein percentage, Milk protein yield	Calving ease, Stillbirth, Gestation length, Scrotal circumference	Residual feed intake Tick resistance	Heel depth Skin coat and texture

15	7	2200000	2299999	73	58	47	Milk beta-caesin percentage, Milk fat yield, Milk fat percentage	Scrotal circumference,	Tick resistance	Skin coat and texture
16	7	12000000	12099999	80	53	40	Milk beta-caesin percentage	Calving ease, Stillbirth, Gestation length, Scrotal circumference	Residual feed intake, Tick resistance	Heel depth
17	7	2300000	2399999	72	49	45	Milk beta-caesin percentage, Milk fat yield, Milk fat percentage	Scrotal circumference	Tick resistance	Skin coat and texture
18	8	110300000	110399999	91	58	51				
19	8	113600000	113699999	82	52	45				
20	8	82400000	82499999	81	50	45	Milk composition		Cold tolerance	
21	8	111600000	111699999	77	46	43				
22	10	13600000	13699999	80	50	49	Milk composition Milk protein percentage	Ovulation rate, Calving ease	Stature	
25	11	24000000	24099999	77	58	41	Milk composition Milk protein percentage	Sperm motility	Residual feed intake, Stature	

23	11	26100000	26199999	78	53	56	Milk composition Milk protein percentage	Sperm motility	Residual feed intake, Tick resistance	
24	11	7400000	7499999	90	48	47	Milk protein percentage	Sperm motility	Residual feed intake,	
28	12	70000000	70099999	184	106	95	Milk composition			
26	12	70200000	70299999	383	97	100	Milk composition			
27	12	70300000	70399999	408	71	58				
30	12	69200000	69299999	138	62	42	Milk fat yield			
31	12	12000000	12099999	80	53	47	Milk yield, Milk protein percentage			
29	12	70100000	70199999	213	46	41	Milk lactose content			
32	12	14900000	14999999	75	45	49	Milk fat yield, Milk protein percentage, Milk yield			
33	13	67700000	67799999	84	61	62	Milk composition	Daughter pregnancy rate, Conception rate		
34	14	22900000	22999999	70	50	45	Milk protein percentage, Milk protein yield, Milk yield	Heifer pregnancy, Age at first calving, Calving ease	Tick resistance, Residual feed intake Stature	Skin coat and texture

35	16	7600000	7699999	100	71	46		Calving ease	PCVI minus PCVM	
36	16	76200000	76299999	92	45	41				
39	17	54500000	54599999	76	66	62	Milk composition			
37	17	70100000	70199999	213	46	41				
38	17	43900000	43999999	154	46	49	Milk composition and Milk protein percentage	Scrotal circumference	Tick resistance	Skin coat and texture
40	18	3900000	3999999	71	64	73	Milk composition	Calving index		
41	19	5300000	5399999	141	45	40	Milk fat yield, Milk protein percentage, Milk composition	Scrotal circumference, Heifer pregnancy		
42	20	70000000	70099999	184	106	49	Milk composition, Milk protein percentage	Heifer pregnancy	Digital dermatitis	Foot and Leg conformation
43	21	24000000	24099999	77	58	67	Milk fat yield	Percentage abnormal sperm		Height
44	21	15200000	15299999	77	51	43	Milk yield	Stillbirth, Gestation length		
46	24	33100000	33199999	71	56	41	Milk composition Milk fat yield, Milk protein yield	Fertility index, Calving ease	Gastrointestinal nematode burden	

45	24	4400000	4499999	87	51	42		Calving ease		
49	25	12600000	12699999	72	56	57	Milk composition		Aggressive behaviour	
48	25	7500000	7599999	84	54	51	Milk fat yield		Residual feed intake	
47	25	9100000	9199999	88	52	53	Milk fat yield, Milk protein percentage,	Dystocia	Residual feed intake, PCV variance	
50	25	15800000	15899999	90	49	41		Calving ease	Tick resistance	Structural soundness Skin coat and texture
51	25	25500000	25599999	72	45	42	Milk yield, Milk composition	Calving ease (maternal), Sperm average path velocity	Gastrointestinal nematode burden	
52	26	28500000	28599999	110	46	52	Milk composition	Gestation length, Twinning	Residual feed intake	Structural soundness, Height (mature)
53	27	9600000	9699999	83	45	40	Milk protein yield, Milk fat percentage	Dystocia	Tick resistance, Gastrointestinal nematode burden	Skin coat and texture

54	29	8200000	8299999	95	54	53		Scrotal circumference	Paired testes volume, Residual feed intake	Structural soundness, Foot angle
55	29	16600000	16699999	81	49	59				Skin coat
56	29	2300000	2399999	72	49	49				
57	X	37100000	37199999	118	49	45				
58	X	88400000	88499999	77	46	42		Scrotal circumference, Age at puberty		
59	X	88400000	88499999	77	46	42		Scrotal circumference, Age at puberty		

Table 4.6: Chromosomal location of identified QTL associated with milk, reproduction, and adaptation traits in high altitude bovines

Chr.	Start	End	Milk traits	Reproduction traits	Adaptation traits
2	3900000	3999999	Milk fat yield	Age at puberty	Lung percentage
7	7600000	7699999	Milk protein %, Milk protein yield	Calving ease, Scrotal circumference	Residual feed intake, Tick resistance, Cold tolerance
7	2200000	2299999	Milk protein %, Milk fat yield, Milk fat %	Scrotal circumference	Tick resistance, Cold tolerance
7	12000000	12099999	Milk protein %	Calving ease, Scrotal circumference	Residual feed intake, Tick resistance, Cold tolerance
7	2300000	2399999	Milk fat yield, Milk fat %, Milk protein %	Scrotal circumference,	Tick resistance, Dry matter intake
10	13600000	13699999	Milk protein %	Ovulation rate, Calving ease	Structural soundness
11	26100000	26199999	Milk palmitic acid content, Milk protein %	Sperm motility	Residual feed intake, tick resistance
11	7400000	7499999	Milk protein %	Sperm motility	Residual feed intake, Feed conversion ratio
14	22900000	22999999	Milk protein %, Milk protein, yield, Milk yield	Calving ease	Dry matter intake, Residual feed intake
17	43900000	43999999	Milk palmitoleic acid content, Milk protein %	Scrotal circumference	Tick resistance
20	70000000	70099999	Milk tridecylic acid content, Milk protein %	Fertility index	Digital dermatitis, Stature
24	33100000	33199999	Milk fat yield, Milk protein yield	Fertility index, Calving ease	Feed conversion ratio
25	9100000	9199999	Milk fat yield, Milk protein %	Calving ease	Residual feed intake, PCV variance
25	25500000	25599999	Milk yield, Milk composition	Sperm average path velocity	Structural soundness
26	28500000	28599999	Milk composition	Gestation length, Twinning	Residual feed intake
27	9600000	9699999	Milk protein yield, Milk fat %	Calving ease	Tick resistance

4.5 Individual selection signatures in high altitude bovines:

Selection signatures are genetic footprints which aid in visualizing the selection process and evolutionary links occurring in a population. Individual selection signatures in high altitude bovines found by combination of SweeD and iHS methods. Number of positive selection signatures along with the number of annotated genes in all three high altitude bovines (Table.4.7). Top 5 genes of all three bovines from both methods in Table 4.8

Table 4.7: **Selection signatures and number of gene annotated in high altitude bovines using *SweeD* and *iHS* methods**

Method	Ladakhi cattle		Ladakhi yak		Yak X cattle hybrid	
	Selection signature	No. of gene annotated	Selection signature	No. of gene annotated	Selection signature	No. of gene annotated
SFS (SweeD)	6346	768	2285	544	2885	534
LD (iHS)	4985	921	1168	342	1945	423
SweeD + iHS		131		78		67

(SFS-Site Frequency Spectrum; LD-Linkage Disequilibirium)

Table 4.8*:* **Top 5 selected genes in high altitude bovines**

Bovines	Gene	SweeD (CLR)	iHS value
Ladakhi cattle	KCNH1	34.6311	5.8059
	DPP6	23.5919	4.9595
	ADAMTSL3	16.6140	3.8105
	RAPGEF5	11.5747	4.4319
	PIEZO2	10.0400	3.9785
Ladakhi yak	OTOL1	52.3605	21.4149
	TMCC3	37.8570	15.7137
	HSD17B2	24.6955	13.1686
	FOXN3	31.5687	10.7068
	LOC109561109	21.6091	9.08617

Yak X cattle hybrid	KLK1	39.6817	14.5842
	RYR2	28.2334	17.6668
	LOC109561556	23.0684	12.6351
	NRXN3	19.3820	11.4068
	LOC109577691	17.9871	11.5602

4.6 Gene annotation and ontology of selected genes in high altitude bovines:

Selected genes from both methods in Ladakhi cattle (131), Ladakhi yak (78) and the yak cattle hybrid (67) were used for further downstream gene ontology analysis to find its biological role and pathways, as described in Table 4.9 and 4.10, respectively.

Table: 4.9: Biological process of selected genes in high altitude bovines

Bovine Group	GO Term	Count	p Value	Genes
Ladakhi cattle	Cell adhesion	6	0.0029	CNTN5, NRXN3, CTNNA3, NCAM2, CTNNA2, ITGA9
	Intracellular signal transduction	5	0.0071	GUCY1A2, PRKCB, RPS6KA2, AKT3, MYZAP
	Potassium ion transmembrane transport	3	0.0323	KCNMA1, KCNAB1, KCNH1
	Cellular calcium ion homeostasis	3	0.0049	PRKCB, ATP13A5, SLC8A1
	Response to hypoxia	5	0.0064	PRKCB, KCNMA1, PML
Ladakhi Yak	Signal transduction	5	0.0355	STARD13, PLPPR1, PLPPR4, PDE6A, ANK3
	Negative regulation of protein catabolic process	3	0.0039	GRIN2A, NOS2, EGFR
	Response to hypoxia	3	0.0307	RYR2, NOS2, HIF1A
	Axon guidance	3	0.0432	ROBO2, ANK3, EFNA5
	Response to redox state	2	0.0223	RYR2, CLOCK
Yak X cattle hybrid	Positive regulation of transcription	4	0.0499	MECOM, ROR2, HIF1A, PPARGC1A
	Response to muscle activity	3	0.0004	RYR2, HIF1A, PPARGC1A
	Cerebral cortex development	3	0.0061	ASPM, FAT4, HIF1A
	Cell proliferation	3	0.0808	APC, COL4A3, NANOG
	Response to hypoxia	2	0.0235	RYR2, HIF1A

The genes under putative selection that were identified in this study accounting for a variety of biological roles such as signal transduction, cellular ion homeostasis, ion transport etc. among this *Response to hypoxia* found in all three high altitude bovines, is an evolutionary adaptation process which ensures the optimal oxygenation of cells. Studies by Friedrich and Wiener., (2020); Iqbal et al., (2019); Graham and McCracken., (2019); Wu et al., (2018) and Yang et al., (2016) reported that long term exposure to hypoxic conditions results in formation of an altered physiological structure for better oxygen delivery, metabolic and cellular homeostasis in livestock such as sheep, goats, cattle, and yaks.

In this study, there was a different set of genes between Ladakhi cattle and Ladakhi yak that regulates response to hypoxia function. The PRKCB, KCNMA1, PML genes were involved in hypoxic response of Ladakhi cattle whereas RYR2, NOS2 and HIF1A genes are associated with the same biological role in Ladakhi yak (Table 4.9). These differences could be due to species variation. KCNMA1 gene acts as an important gene for hypoxic response by modulating cellular adaptation in cardiac, smooth, and skeletal muscles that assist in cardiovascular conduction and cellular energy homeostasis in mountain deer and goats (Schweizer et al., 2019; O'Brien et al., 2020). It can be assumed that the KCNMA1 might behave the same in Ladakhi cattle. The HIF1A is the most commonly reported gene, which triggers the hypoxic related modulation in various of high-altitude livestock. It acts initiates the cellular survival, proliferation, oxygen transportation, cardiac regulation and energy homeostasis (Xiong et al., 2015; Lu et al.,2018). This finding suggests that each species has their own mechanism to adapt.

The selected genes were involved in signaling pathways (Table 4.10), cell to cell communication, calcium ion transportation, cellular metabolism, hormonal regulation, and organogenesis triggered hypoxia. Selection signatures accounting for Hypoxia Induced factor and angiogenesis pathways in all bovine groups are vital for survival under chronic hypoxia conditions, similar observations were reported in cashmere goats (Song et al., 2016), Tibetan sheep (Yang et al., 2016) and Ethiopian cattle (Edea et al., 2013). Selection signatures of high-altitude bovines in this study were also found to be associated with energy metabolism, oxidation reactions, stress response and influencing body morphology, similar reports were observed in various other studies related to high altitude livestock (Edea et al., 2014; Yang et al., 2016; Gorkhali et al., 2016)

Table 4.10: Top 5 molecular pathways of selected genes in high altitude bovines

Bovine Species	Molecular pathway	Count	p value	Genes
Ladakhi cattle	Cell adhesion molecules (CAMs)	5	0.0094	CNTNAP2, NRXN3, ITGA8, NCAM2, ITGA9
	cGMP-PKG signaling pathway	5	0.0105	GUCY1A2, AKT3, KCNMA1, PDE3A, SLC8A1
	Purine metabolism	5	0.0148	GUCY1A2, PDE1C, PDE3A, FHIT, AK9
	MAPK signaling pathway	5	0.0485	NTRK2, PRKCB, RPS6KA2, AKT3, MAP3K4
	Neurotrophin signaling pathway	4	0.0299	NTRK2, RIPK2, RPS6KA2, AKT3
Ladakhi yak	PI3K-Akt signaling pathway	10	0.0000	FGF14, ITGB5, PPP2R2C, EGF, PPP2R2B, COL5A2, ITGB6, FGF13, EFNA5, EGFR
	Regulation of actin cytoskeleton	6	0.0010	FGF14, ITGB5, EGF, ITGB6, FGF13, EGFR
	Rap1 signaling pathway	6	0.0010	GRIN2A, FGF14, EGF, FGF13, EFNA5, EGFR
	Calcium signaling pathway	5	0.0049	RYR2, GRIN2A, NOS2, RYR3, EGFR
	HIF-1 signaling pathway	4	0.0052	NOS2, EGF, HIF1A, EGFR
Yak X cattle hybrid	Angiogenesis	5	0.0000	APC, HIF1A, RYR2, GJA1, RYR3
	Hypoxia response	3	0.0228	HIF1A, RYR2, CHRNA7
	Wnt signaling pathway	3	0.0146	FAT4, APC, ANK3
	Cadherin signaling pathway	2	0.0040	FAT4, APC
	Integrin signaling pathway	2	0.0234	COL4A3, COL6A57

4.7 Co selected genes in high altitude bovines:

Co-selected genes of these three groups identified using the XP-EHH method using pairwise comparison, those co-selected genes annotated with its biological role using previously reported studies (Table 4.11). Notably, the DGKG. which regulates epidermal growth factor receptors, tyrosine kinase receptors and also reduces oxidative stress in high altitude conditions (Ishisaka and Hara.,2014; Zhou et al., 2014). The PBX4 have been positively selected against chronic mountain sickness and influences spermatogenesis (Wagner et al., 2001; Zhou et al., 2014).

Table 4.11: Co selected genes of high-altitude bovines along with their biological role

Gene	Selected position on chr.	XP-EHH	Biological Role	Reference
TMEM45A (Transmembrane protein 45A)	chr1:46187794	2.4557	Epidermal morphogenesis, keratinization and its maintenance.	Hayez et al., 2016
DGKG (Diacylglycerol kinase gamma)	Chr1: 82942668, 82942669, 82942670	2.3843	Regulates epidermal growth factor receptor and tyrosine kinase receptors. Oxidative stress response in high altitude.	Ishisaka and Hara., 2014 Zhou et al., 2014
PPARA (Peroxisome Proliferator-Activated Receptor α)	Chr5:123166373	3.4345	Regulation of fatty acid oxidation and glucose. Increase oxygen utilization in high altitudes	Tian et al.,2020 O'Brien et al., 2020
PBX4 (PBX Homeobox 4)	Chr7:3658198, 3658201	2.6287	Positively selected against chronic mountain sickness. Spermatogenesis.	Zhou et al., 2014 Wagner et al., 2001.
PTPRD (Protein tyrosine phosphatase receptor type D)	Chr8: 38246110	2.2113	Anti-cancer agent against cutaneous squamous cell carcinoma,melanoma and lung cancer	Solomon et al., 2008. Kohno et al., 2010 Lambert et al., 2012
FAM184A (Family with sequence similarity 184 member A)	Chr9:33859105, 33859122	2.6503	Anti-cancer property	Liu et al., 2019
FNDC1 (Fibronectin type III domain-containing protein 1)	Chr9:99370537	2.6545	Survival of cardiomyocytes under hypoxia	Sato et al., 2009. Liu et al.,2019
UNC13C (Unc13 Homolog C/Munc 13-3)	Chr10: 56761935	2.8915	Protects neurons and maintain neuroplasticity	Yang et al., 2007
GALNT16 (Polypeptide N-Acetyl galactosaminyl transferase 16)	Chr10:83290369	2.1986	Association with milk protein and fat	Gao et al., 2017
PDZRN3 (PDZ Domain Containing Ring Finger 3)	Chr22: 29058364, 29058378, 29058404	2.8296	Developmental processes	Honda and Inui., 2015.

4.8 Gene interaction network in high altitude bovines:

Common selected genes among high altitude bovines were sorted out and used to construct interactive networks associated with a variety of adaptive functions that aids in maintain cellular and metabolic homeostasis, protection against the adverse conditions and alter morphogenesis which comprehensively improves the survivability in high altitude conditions

4.8.1 Adaptive hypoxia regulation in high altitude bovines

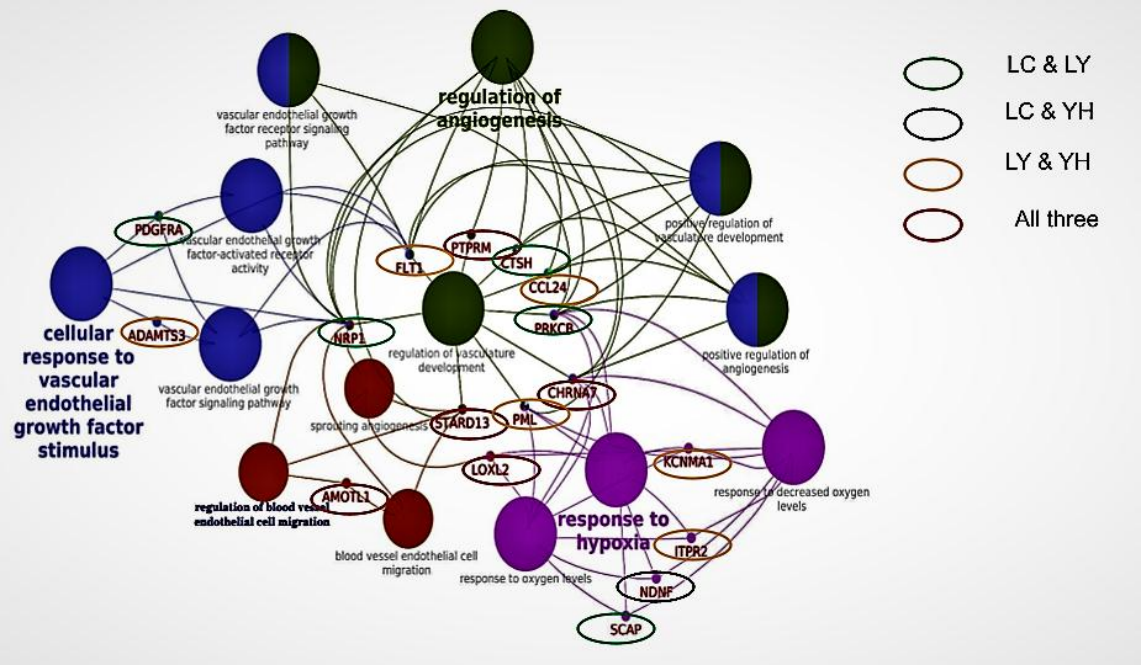
The selected genes were used to construct a gene-to-gene network of hypoxic regulation consisting of three major adaptive mechanisms such as altering of blood cells size and volume, blood vessel size and cardiac chambers for better oxygen transportation to aid cellular metabolism.

Notably selected genes such as protein tyrosine phosphatase receptor type M (PTPRM), star related lipid transfer domain containing 13 (STARD13), cholinergic receptor nicotinic alpha 7 subunit (CHRNA7), lysyl oxidase like 2 (LOXL2) and angiomotin-like protein 2 (AMOTL2) were found common among all three bovines in Figure. 4.5.

The PTPRM is an anti-cancer gene that has a regulatory role in cell-to-cell communication, cellular proliferation and maintaining cellular integrity in vascular endothelial cells (Sun et al., 2012). Studies on high altitude chickens by Tian et al., (2020) showed importance of PTPRM gene for hypoxic adaptation in terms of enlarging blood vessels. The STARD13, a vital gene that modulate the lung epithelial cells and has anticancer property (Sun et al., 2018; Basak et al., 2018), Sun et al., (2018) found the STARD13 plays a crucial role in reptiles to sustain in high altitude hypoxic conditions via maintaining the tissue morphogenesis of lungs.

The CHRNA7 is a cholinergic receptor which tunes and protect neural cells (Lu et al., 2004; Schmidt-Kastner et al., 2006), Heeschen et al., (2002); Schmidt-Kastner et al., (2006) studies showed CHRNA7 gene is upregulated in endothelial cells during proliferation, by hypoxia in-vitro, and by ischemia in-vivo which suggests that it has a vital role in angiogenesis. Recent study by Qu et al., (2015) on Tibetan Yak showed CHRNA7 gene involved in Hypoxia induced Factor, Calcium signaling and Angiogenesis pathways.

Figure 4.5: Gene to Gene interaction network of adaptive hypoxic regulation



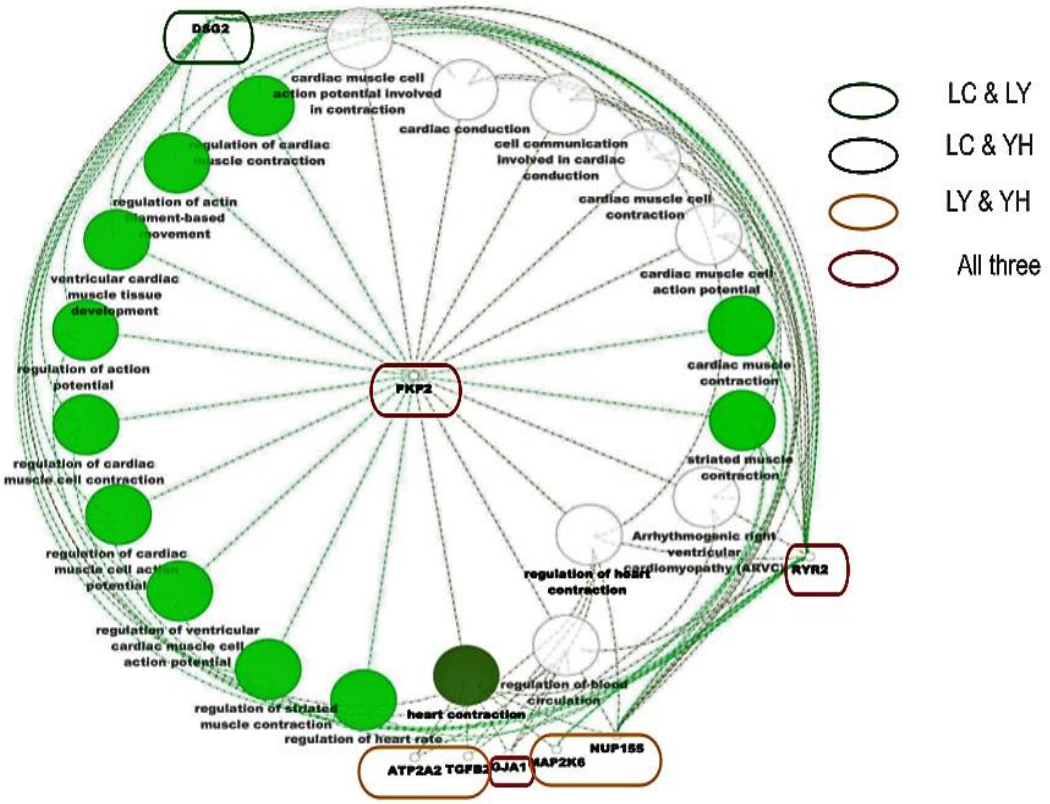
(*Green* - Common between Ladakhi cattle and Ladakhi yak, *Black* - Common between Ladakhi cattle and Yak hybrid, *Yellow* - Common between Ladakhi yak and Yak hybrid, *Red* - Common to all groups)

The LOXL2 possesses an important role in stabilization and maintaining the integrity of collagen and elastin fibrils in smooth muscles (Saleem and Rajput., 2020). Li et al., (2021) found structural development of lung tissues and intensity of expression in trachea's elastic fibers that provides a strong retraction force during the exchange air between the outside atmosphere and the blood in the lungs of Tibetan yaks. AMOTL2 mediates angiostatin migration in endothelial cells and blood vessel morphogenesis (Aase et al.,2007; Li et al., 2012), no prior reports on AMOTL2 in livestock but it was detailed in humans, mice, and zebrafish (Wang et al., 2011; Huang et al., 2018).

4.8.2 Adaptive cardiac regulation in high altitude bovines

High altitude conditions causes pulmonary arterial hypertension that develops dilatory heart failure (Bärtsch and Gibbs., 2007; Naeije., 2010; Netzer et al., 2013) to overcome this adverse effect on long term exposure an adaptive cardiac regulation occurs as a natural evolutionary response by selected genes which controls the overall cardiac morphogenesis, heart rate, contraction, action potentials, blood flow and cardiac output in high altitude bovines, thus aids to counter the hypoxic situation (Figure 4.6).

Figure 4.6: **Gene to Gene interaction network of adaptive cardiac regulation**



(*Green* - Common between Ladakhi cattle and Ladakhi yak, *Black* - Common between Ladakhi cattle and Yak hybrid, *Yellow* - Common between Ladakhi yak and Yak hybrid, *Red* - Common to all groups)

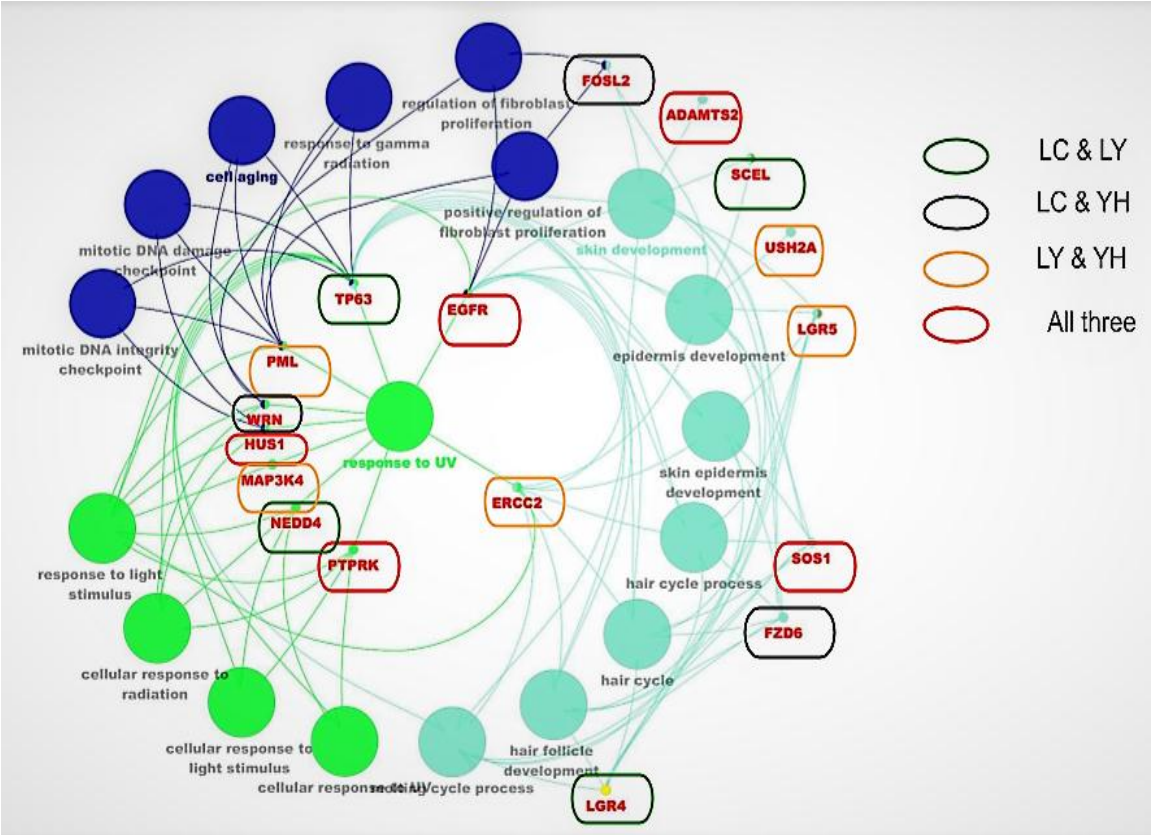
In this study, *plakophilin 2* (PKP2), *ryanodine receptor 2* (RYR2) and *gap junction alpha-1 protein* (GJA1) were found to be common in all high-altitude bovines (Figure 4.5). These genes are found to be involved in cardiomyopathy i.e enlargement of heart, alters the cardiac rhythm, controls calcium signaling and blood circulation in humans (Hwang et al., 2011; Cerrone et al.,2017; Bround et al., 2012).

4.8.3 Adaptive cutaneous regulation in high altitude bovines

High altitude conditions such as low humidity, hypoxia, severe cold and high UV radiation exposure tends to cause variety of cutaneous related problems notably xerosis (dry skin), frostbite, pruritus, burns, dermatitis and cutaneous cancer in humans (Singh., 2017; Pons- Guiraud., 2007; Hawk et al., 2010). But the livestock and animals of high altitude adapt to these conditions via altered genetic architecture causing hyper pigmentations, growth of dense hair follicles, less or no sweat glands and dense cutaneous layers (Figure 4.7)

In this study, the genes that are selected to cause the cutaneous adaptation were mapped as shown in Figure 4.7. The protein tyrosine phosphatase receptor type K (PTPRK), HUS1 checkpoint clamp component (HUS1), SOS Ras/Rac guanine nucleotide exchange factor 1 (SOS1), epidermal growth factor Receptor (EGFR) and a disintegrin and metalloproteinase with thrombospondin motifs 2 (ADAMTS2) were found common among all three bovine group of high altitude. These genes were not known commonly in livestock, *PTPRK*, *HUS1* and *SOS1* were positively selected tumor suppressor genes which present in epithelial cells, collagen fibrils of skin and hair follicles in humans and mice (Chang et al., 2020; Zhou et al., 2019; Cai et al., 2019). The researchers found these genes were influencing the coat pigmentation, hair growth and cellular response to radiation in the high lander of Tibet and Nepal regions (Beall et al., 2007; Simonson et al., 2010 and Cole et al., 2017).

Figure 4.7: Gene to Gene interaction network of adaptive cutaneous regulation



(*Green* - Common between Ladakhi cattle and Ladakhi yak, *Black* - Common between Ladakhi cattle and Yak hybrid, *Yellow* - Common between Ladakhi yak and Yak hybrid, *Red* - Common to all groups)

EGFR regulates various keratinocyte functions such as proliferation, adhesion and migration, survival, and differentiation. Some studies showed EGFR signaling governs cell fate of keratinocytes transitioning from the basal proliferative compartment to the differentiating supra-basal layers in skin (Jost et al., 2000; Pastore et al., 2008; Tran et al., 2010) in humans. But the same gene involved in maintain of uterine health and regulates fertility in cattle (Kliem et al.,1998; Katagiri and Takahashi., 2004; Allen et al., 2017; Takatsu et al., 2018), sheep (Gharib-Hamrouche et al.,1995; Tamada et al., 2002), goats (Gall et al., 2004; Tibary et al., 2005; Paramio and Izquierdo et al., 2014) and yaks (Pan et al., 2014; Zhou et al., 2020). Li et al.,2021 found ADAMTS2 gene is important for skin integrity in yaks, Similar findings were found by Zhou et al., (2012) in sheep, Lühken et al., (2012) in goats and Jaffey et al., (2019) in dogs.

4.8.4 Adaptive energy metabolism in high altitude bovines

In the absence of sufficient oxygen, energy production from oxidative metabolism may be reduced. Additionally, if oxidative metabolism proceeds in hypoxic conditions it causes accumulation of reactive oxidative intermediates in mitochondria that results in both energy depletion and oxidative stress at cellular level (Scortegagna et al., 2003; Ge et al., 2015). An evolutionary metabolic change at cell level is induced by hypoxia, causing a variety of alternate energy metabolism notably fatty acid oxidation (Holden et al., 1995; Koves et al., 2008; Rankin et al., 2009; Simonson et al., 2010), anaerobic glucose metabolism (Semenza., 2009; Kelly et al., 2010; Majmundar et al., 2010), gluconeogenesis (Rankin et al., 2008; Pescador et al., 2010) and glycolysis (Kim et al, 2006 ; Papandreou et al.,2006 ; Simonson et al., 2010) in order to maintain energy production.

In this study, a group of positive selected genes were found in high altitude bovines that tend to influence energy cellular metabolism (Figure 4.7). The AKT serine/threonine kinase 3 (AKT3), insulin receptor substrate 2 (IRS2), parathyroid hormone (PTH), insulin like growth factor binding protein 5 (IGFBP5), peroxisome proliferator activated receptor alpha (PPARA) and early growth response protein 1 (EGR) genes were found common among all three bovine groups (Figure 4.8).

The PPARA gene regulates lipid metabolism in the liver and skeletal muscle, which aids in glucose homeostasis. The transcriptional regulation of genes involved in peroxisomal and mitochondrial -oxidation pathways, fatty acid absorption, and

triglyceride catabolism regulates intracellular lipid and carbohydrate metabolism (Lefebvre et al., 2006; Tyagi et al., 2011). In yak, Qin et al. (2015) discovered its relevance in skeletal muscle lipid metabolism. PPARA gene affects systemic energy metabolism in Tibetan Plateau livestock and people, according to Wei et al., (2016); Jia et al., (2016); Friedrich and Wiener, (2020).

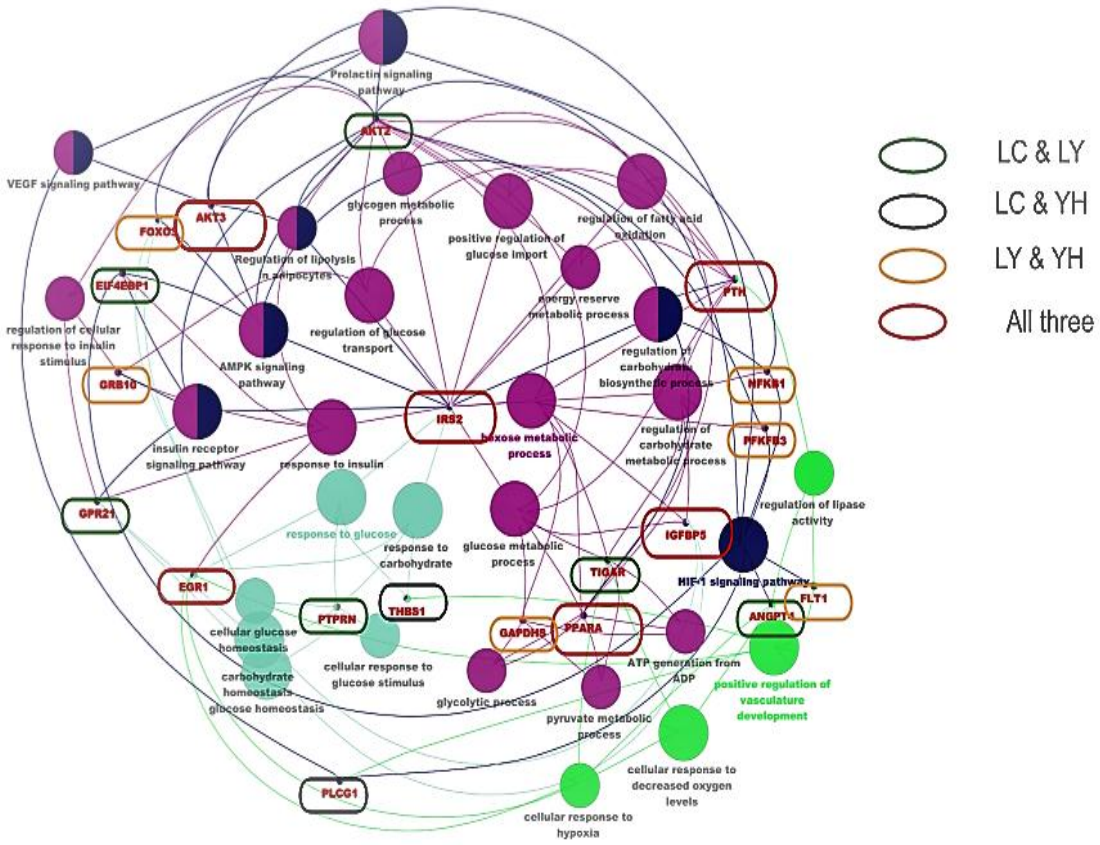


Figure 4.8*:* Gene to Gene interaction network of adaptive energy metabolism

(*Green* - Common between Ladakhi cattle and Ladakhi yak, *Black* - Common between Ladakhi cattle and Yak hybrid, *Yellow* - Common between Ladakhi yak and Yak hybrid, *Red* - Common to all groups)

The EGR1 gene is thought to be involved in the generation of gluconeogenesis and cholesterol in the liver (Shen et al., 2015). It regulates important metabolic processes such adipocyte insulin resistance, energy storage, and insulin production (Yu et al., 2011; Muller et al., 2012; Zhang et al., 2013). According to Zhang et al., (2017), the EGR1 gene regulates gluconeogenesis in skeletal muscles under hypoxia, thus it can be predicted that the EGR1 gene is operating in a similar way in yaks also.

During hypoxic conditions in humans, the AKT3, IGFBP5 and IRS2 genes have a feedback regulatory mechanism that regulates the phosphatidylinositol-3 kinase (PI-3K)/Akt pathway and insulin growth factors (IGF) to aid in angiogenesis and change energy metabolism (Lum et al., 2007; Nam et al., 2011; White., 2014). However, similar reports have not been discovered in animals, we believe these genes may affect adaptive energy metabolism.

4.9 Identification of genome wide co-selective regions between Ladakhi and high producing cattle:

To understand the genomic similarities among Ladakhi cattle with high producing *Bos indicus* (Sahiwal) and *Bos taurus* (Holstein Friesian and Jersey) cattle breeds, co-selected regions associated with strong positive selection signatures in the genome were identified.

4.9.1 Selected markers and shared selected regions between Ladakhi and high producing cattle:

Initially, the genomic markers of Ladakhi cattle were matched with that of high producing cattle to find common markers between them. The Holstein Friesian has a greater proportion (34%) of common markers (161931 of 483635 markers) followed by Sahiwal (28.68 %, 59130 of 206172 markers) and Jersey (24. 82 %, 113947 of 459093 markers) with Ladakhi cattle. However, combine selection sweep analysis using XP-EHH and RsB methods showed that the Sahiwal had a greater number of high significant ($p < 0.05$) co selection signatures (2055) and shared selection regions (539) with that of Ladakhi cattle followed by Holstein Friesian (1645 and 286) and Jersey (1208 and 215) (Table 4.12). The greater proportion of common positive selection signature between the Ladakhi and the Sahiwal indicated their evolutionary closeness and suggests they share some closure genetic similarities compared to Ladakhi and the *Bos taurus* (HF and Jersey) cattle.

4.9.2 QTL mapping of shared selected regions between Ladakhi and high producing cattle

Shared selection regions with greater than 2 score in both XP-EHH and RsB were filtered out and considered as highly significant positively selected shared regions between Ladakhi cattle and high producing cattle. Sahiwal shared 52 highly significant positively selected regions followed by Holstein Friesian (32) and Jersey (29) with

Ladakhi cattle. These positively selected shared regions were further mapped for the presence of QTL and classified trait wise. For milk, reproduction, and adaptation traits, respectively, a total of 40, 25 and 28 shared selected regions between Ladakhi with Sahiwal, 23, 17 and 11 between Ladakhi and HF and 19, 15 and 17 between Ladakhi and Jersey.

Table 4.12: Proportion of significant selected markers and shared selected regions between Ladakhi and high producing cattle

	Ladakhi vs. Sahiwal		**Ladakhi vs. HF**		**Ladakhi vs. Jersey**	
Method	**Significant markers (p <0.05)**	**Shared selected regions**	**Significant markers (p <0.05)**	**Shared selected regions**	**Significant markers (p <0.05)**	**Shared selected regions**
XP-EHH	3020	605	2425	460	1891	340
RsB	3290	653	2532	476	1983	347
Both (XP-EHH+RsB)	2055	539	1645	286	1208	215

Shared selected regions common for all three traits between Ladakhi and Sahiwal cattle are described in Table 4. 13. Notably, chr 7: 7100000-7200000 region associated with 4 selected markers was associated with a vital adaptive trait of cold tolerance to high altitude conditions and chr22: 42600000-42700000 region has more selected markers (5) and found have influence over the stature and skeletal morphology associated with physical fitness. Such selected regions between Ladakhi and HF are described in Table 4.14, where in, chr 10: 25200000-25300000 region associated greater selected markers (7) has been found to be associated with adaptive trait of subcutaneous fat deposition to tackle energy deficit conditions and also has influence over traits such as stature and calving ease associated with physical fitness and reproduction respectively. Between Ladakhi and Jersey cattle, the shared 11900000-12000000 region of chromosome 7 is associated with greater selected markers (17) which influences calving ease a adaptive trait for pregancy difficulties and 20200000-20300000 shared region of chromosome 20 is important for adaptive trait of hematocrit for blood size and blood volume.

Table 4.13: **Shared selected regions common for milk, reproduction and adaptation traits between Ladakhi cattle and Sahiwal**

Chr.	Start	End	No. of selected markers	Milk	Reproduction	Adaptation
4	17200000	17300000	3	Milk protein percentage	Male fertility	Residual feed intake
5	32300000	32400000	3	Milk protein %, Milk yield, Fat yield	Twinning, Calving ease	Tick resistance, Skin coat
7	7100000	7200000	3	Milk protein %	Male fertility, Calving ease	Tick resistance, Residual feed intake, Cold tolerance
8	27900000	28000000	2	Milk yield	Calving ease	Residual feed intake
10	9900000	10000000	4	Milk yield	Calving ease	Foot and leg conformation
10	10700000	10800000	3	Milk protein yield	Calving ease	Structural soundness
10	15100000	15200000	3	Protein %, Fat yield	Calving ease	Immunoglobulin G level
12	37000000	37100000	4	Fat yield	Fertility	Structural soundness
22	42600000	42700000	5	Protein %	Calf Size,	Stature
23	19000000	19100000	3	Fat yield, protein %	Sperm motility	Tick resistance
23	27000000	27100000	3	Milk yield, Fat yield, Protein yield	Male fertility, Calving ease	Tick resistance, Skin coat

Table 4.14: **Shared Selected regions common for milk, reproduction, and adaptation traits among Ladakhi cattle and HF**

Chr.	Start	End	No. of Selected markers	Milk	Reproduction	Adaptation
7	8300000	8400000	5	Protein %, Protein Yield	Calving ease	Residual feed intake, Tick resistance
10	25200000	25300000	7	Fat yield, Protein %, Protein yield	Calf size, Calving ease	Stature, Subcutaneous deposition
18	28700000	28800000	2	Fat yield, Milk yield	Fertility index	Residual feed intake
23	28600000	28700000	4	Protein %, Protein yield	Male fertility	Residual feed intake, Tick resistance

Table 4.15: **Shared Selected regions common for milk, reproduction, and adaptation traits among Ladakhi cattle and Jersey**

Chr.	Start	End	No. of Selected markers	Milk	Reproduction	Adaptation
2	29100000	29200000	4	Protein %, Milk yield, Protein yield, Fat %	Fertility, Calving ease,	Hematocrit (Cell volume)
7	11900000	12000000	17	Protein %, Protein yield	Calving ease	Residual feed intake
10	36900000	37000000	3	Fat yield, Protein %	Calving ease	Tick resistance, Subcutaneous deposition
14	13500000	13600000	16	Milk yield	Calving ease, Twinning	Stature, residual Feed intake, Tick resistance
15	46200000	46300000	6	Milk yield	Calving ease, Male Fertility	Structural soundness
20	20200000	20300000	8	Fat %, Protein %, Protein yield	Sperm motility, calving index, Calving ease, Calf size	Hematocrit (RBC size, Volume), Stature, Immunoglobulin level
28	18100000	182000000	2	Milk yield	Male fertility	Structural soundness

4.9.3 Shared selected regions between Ladakhi cattle and high producing cattle related to adaptation and reproduction traits:

Co-selection signatures are indications of shared genetic regions that help to access the genomic similarity and the evolutionary linakge between two populations (Gutiérrez-Gil et al., 2015; Ghoreishifar et al., 2020). To relate and compare the genetic compatibility between two population, the QTL analysis of co selected regions were carried out to map the regions associated with metabolic homeostasis or other general traits such as disease resistance and behavior (Gutiérrez-Gil et al., 2015; Randhawa et al., 2016; Ben-Jemaa et al., 2020). Genetic compatibility between Ladakhi and high producing cattle can be better understood by enumerating the proportion of adaptive and reproductive traits vital for high altitude survival. Major traits such as external morphology, fertility, cold tolerance, residual feed intake, hematocrit and calving ease were sorted and enumerated between each comparison (Table 4.16).

In this study, the Ladakhi and the Sahiwal cattle comparison has a greater number of shared regions for external morphology (12) followed by the Jersey (7) and the Holstein Friesian (4), which shows that the Sahiwal cattle might develop an identical external morphology as of the Ladakhi cattle that could withstand and thrive in the steep high altitude of Ladakh region. External morphology related traits such as skeletal conformation, stature, leg and feet conformation helps the humans and animals in high altitude to well adapt in the mountainous conditions, which strengthens locomotion in steep mountains. Similarly, skin coat and skin texture related traits help the stronger cutaneous coat that could withstand higher radiations of high altitude (Storz et al., 2010; Jeong et al., 2014; Friedrich and Wiener., 2020).

Table 4.16: **Shared selected regions between Ladakhi cattle and high producing cattle related to adaptation and reproduction traits**

Traits	Ladakhi vs. Sahiwal		Ladakhi vs. HF		Ladakhi vs. Jersey	
	No. of putative selected regions	**Region description**	**No of putative selected regions**	**Region description**	**No of putative selected regions**	**Region description**
External Morphology (structural soundness, foot and leg, stature, skin coat)	12	2:8000000-8100000, 2:62000000-62200000, 5:32300000-32400000, 7:7100000-7200000, 7:87100000-87200000, 8:42200000-42300000, 10:9900000-10000000, 10:4300000-4400000, 12:37000000-37100000, 21:15300000-15400000, 22:42600000-42700000, 23:5600000-5700000	4	4:113300000-113400000, 10:25200000-25300000, 12:36800000-36900000, 16:43900000-44000000	7	3:69500000-69700000, 4:113300000-113400000, 13:6600000-6700000, 14: 13500000-13600000 15:46200000-46300000, 28:18100000-18200000, X:133100000-133200000
Fertility	11	1:40000000-40100000, 1:47300000-47400000, 4:17200000-17300000, 7:7100000-7200000, 10:62200000-62300000, 10:70800000-70900000, 12:37000000-37100000,	11	1:131000000-131100000, 1:59900000-60000000, 10:70800000-70900000, 15:3500000-3600000, 16:43900000-44000000,	9	1:130900000-131100000, 2:29100000-29200000 5:117000000-117100000 6:59500000-59600000, 7:11900000-12000000, 15:46200000-46300000, 15:69900000-70000000,

		23:5600000-5700000, 23:19000000-19100000, 24:55600000-55700000, X:105300000-105400000		17:70900000-71100000, 18:28700000-28800000, 21:10200000-10300000, 23:28600000-28700000, 25:34000000-34100000 X:109400000-109500000		20:20200000-20300000, 21:1200000-1300000, 28:18100000-18200000, X:133100000-133200000
Calving ease	8	5:32300000 - 32400000, 8:27900000 - 28000000, 8:42200000 - 42300000, 10: 9900000 - 10000000, 10:10700000-10800000, 10:15100000 - 15200000, 23:27000000 – 27100000	5	6:59500000-59600000, 6:31400000-31500000, 7:8300000-8400000, 10:25200000-25300000, 25:34000000-34100000	7	2:29100000-29200000, 3:54500000-54700000, 5:117000000-117100000, 7:11900000-12000000, 10:36900000-37000000, 13:6600000-6700000, 14:13500000-13600000, 15:46200000-46300000
Cold tolerance	3	7:7100000-7200000, 7:79800000-79900000, 7:87100000-87200000	Nil	Nil	Nil	Nil
Hematocrit (Blood cells volume and size)	2	11:73800000-73900000, 20:17500000-17600000	1	14:79400000:79500000	2	2:29100000-29200000, 20:20200000-20300000

Cold tolerance is also genetically regulated trait, where changes occurring at genome level may alter the phenotype, such as less to no cutaneous sweat glands, thick skin coat and alter basal metabolic rates (Wei et al., 2016; Hu et al., 2019; Ayalew et al., 2021). In this study, Cold tolerance related shared regions were found only in Ladakhi and Sahiwal cattle (3 regions) (Table 4.16), which indicates that these two breeds might have evolved through similar mechanisms to adapt to cold stress at genome level.

Hematology related traits such as hematocrit (HCT), mean cell volume (MCV) and mean corpuscular hemoglobin concentration (MCHC) percentage make change in blood cell size and blood volume that help to normalize the cellular hypoxic conditions and energy dynamics in high altitude conditions (Ding et al., 2014; Lu et al., 2019; Liu et al., 2020). In this study, Ladakhi vs. Sahiwal and Ladakhi vs. Jersey had the same number of shared regions (each 2), (Table 4.16) associated with hematological traits. The findings indicate that both Sahiwal and Jersey breeds might develop similar sorts of blood volume and size.

Reproduction is a major concern to address in cold hypoxic high-altitude condition as these combination stress would cause a variety of abnormalities in structural and hormonal differences, resulting in irregularities in estrus cycle and pregnancy difficulties. To overcome to this, the high-altitude livestock have been selected for improvement on certain traits such as fertility index, ovulation rate, sperm motility and calving ease, under natural adaptation process over the period (Moore et al., 2004., Parraguez et al., 2013; Scheinfeldt et al., 2012; Moore., 2021; Gonzalez-Candia et al., 2021). For reproduction related traits, the Ladakhi with Sahiwal had greater number of shared regions (19), combined fertility and calving ease, followed by Ladakhi with Jersey (16) and Ladakhi with Holstein Friesian (15) (Table 4.16). This result implies that the Sahiwal might have high similarity in reproduction efficiency and calving ease with Ladakhi cattle compared to HF and Jersey.

These traits collectively assist for better survivability and sustainability, especially in high altitude conditions. In our study based on QTL analysis, based on comparison of genetic compatibility among various high producing breeds (Sahiwal, HF and Jersey) with Ladakhi cattle, it can be concluded that Sahiwal has more proximity toward Ladakhi cattle compared to HF and Jersey in terms of adaptation and reproduction traits.

4.9.4 Shared selected regions related to milk trait between Ladakhi cattle and high producing cattle

Milk production is the prime goal of dairy animals, a variety of breeding programs were followed to improve low producing population. Milk composition realted traits protein yield, protein percentage, fat yield and fat percentage is gaining importance in the milk product industries. In this study, the possibility to improve Ladakhi cattle milk production by comparing co-selected shared regions with high producing cattle (Sahiwal, Holstein Friesian and Jersey) for milk production related traits was assessed. Economically important traits such as milk yield, fat percentage, fat yield, protein percentage and protein yield were considered and enumerated in pair wise comparisons of between Ladakhi and Sahiwal, Ladakhi and HF, finally as Ladakhi and Jersey (Table 4.17).

There were some of the important differences observed between the cattle breeds comparisons related to shared selected regions associated with milk protein percentage and milk protein yield, milk fat percentage and milk fat yield traits. The Ladakhi cattle had the greater number of shared regions for these traits with Sahiwal compared to HF and Jersey (Table 4.17). This implies that Ladakhi cattle might have a greater analogy of milk traits with Sahiwal compared. Shared selected regions related to milk yield were found more or less identical among all three comparisons which implies that all high producing cattle might have similar sort effect with Ladakhi cattle.

Table 4.17: **Shared selected regions related to milk trait between Ladakhi and high producing cattle**

	Ladakhi vs. Sahiwal		Ladakhi vs. HF		Ladakhi vs. Jersey	
Traits	**No. of putative selected regions**	**Region description**	**No. of putative selected regions**	**Region description**	**No. of putative selected regions**	**Region description**
Protein percentage and Protein yield	13	1:40000000-40100000, 1:47300000-47400000, 2:8000000-8100000, 5:32300000-32400000, 7:7100000-7200000, 10:62200000-62300000, 10:15100000-15200000, 17:59700000 - 59800000, 17:65900000-66000000, 18-54500000-54600000, 20:17500000-17600000, 23:19000000-19100000, 29:43400000-43500000	10	1:59900000-60000000, 6:31400000-31500000, 7:8300000-8400000, 7:11900000-12000000, 9:34300000-34400000, 10:25200000-25300000, 10:70800000-70900000, 10:77800000-77900000 23:28600000-28700000, 25:34000000-34100000	8	2:29100000-29200000, 6:59500000-59600000, 7:11900000-12000000, 10:36900000-37000000, 11:52600000-52700000, 20:20200000-20300000, 29:12900000-13000000, 29:30900000-31000000
Fat % and Fat yield	17	1:47300000-47400000, 2:62000000-62200000, 2:89200000-89300000, 2:8000000-8100000, 5:32300000-32400000,	9	2:57500000-57600000, 6:59500000-59600000, 6:31400000-31500000, 9:34300000-34400000, 10:25200000-25300000,	4	2:29100000-29200000, 6:59500000-59600000, 10:36900000-37000000, 20:20200000-20300000

		6:116200000-116300000, 10:62200000-62300000, 10:70800000-70900000, 10:75800000-75900000, 10:15100000-15200000, 11:73800000-73900000, 12:37000000-37100000, 15:58600000-58700000, 18:54500000-54600000, 20:17500000-17600000, 23:19000000-19100000 27:2200000-2300000		10:70800000-70900000 10:77800000-77900000, 12:36800000-36900000, 15:48000000-48100000		
Milk yield	8	5:32300000-32400000, 8:27900000-28000000, 10:9900000-10000000, 10:70800000-70900000, 10:10700000-10800000, 11:79800000-79900000, 16:74800000-74900000, 23:27000000-27100000	10	2:57500000-57600000, 6:59500000-59600000, 6:31400000-31500000, 9:34300000-34400000, 10:70800000-70900000, 12:36800000-36900000, 15:48000000-48100000, 18:28700000-28800000, 21:10200000-10300000, 25:34000000-34100000	10	2:29100000-29200000, 3:69500000-69700000, 6:59500000-59600000, 11:52600000-52700000, 14:13500000-13600000, 18:20000000-20100000, 21:1200000-1300000, 28:18100000-18200000, 29:12900000-13000000, 29:30900000-31000000

4.9.5 Common Selection signatures between Ladakhi and high producing cattle

Common selection sweep analysis was carried out between Ladakhi cattle and other high producing breeds, pairwise. The study indicated, as shown by commonly selected regions earlier, that Holstein Friesian had a greater proportion of common markers, however, selection signatures were found more in the Sahiwal cattle. These findings also reach the same conclusion that the Ladakhi cattle might have more evolutionary closeness with Sahiwal compared to the HF and Jersey (Table 4.18). Further, a detailed analysis of co-selected shared regions between Ladakhi and High producing cattle showed Sahiwal has a superior number of selected regions for traits such as external morphology, fertility and calving ease. Another striking observation was the presence of cold tolerance traits related to selected regions only in the Ladakhi and Sahiwal pair comparison, which suggests that there is similar genetic evolution between Ladakhi and Sahiwal cattle for the cold tolerance (Table 4.18)

Table 4.18: Outline of selection sweep after comparison of Ladakhi cattle with high producing breeds

Breed (compared with Ladakhi)	Percentage of common markers	Significant selected sweeps	Significant selected regions	Highly selected regions	Notable similarity with Ladakhi cattle
Sahiwal	28.68% (59130)	2055	539	52	Fertility, Calving ease, Cold tolerance, Structural soundness, Blood cell volume and size
Holstein Friesian	33.48% (161921)	1645	286	32	Fertility, Calving Ease, Subcutaneous fat deposition, Blood cell volume
Jersey	24.82 % (113497)	1208	215	29	Fertility, Calving Ease, Structural soundness, Subcutaneous fat deposition, Blood cell volume

4.10 Genome wide Co-selective regions between Ladakhi yak and cattle

Co-selected regions among Ladakhi yak and cattle (Ladakhi cattle; Sahiwal (*Bos indicus*) and Holstein Frisian; Jersey (*Bos taurus*) were analyzed, using similar combinational approach as of Ladakhi cattle with high producing cattle. Comparative selection sweeps analysis showed a greater number of common markers between Ladakhi cattle and Ladakhi yak (5581); approximately sharing of 1.11 % of total makers, followed by Sahiwal (2744 markers), HF (2553 markers) and Jersey (1972 markers).

A detailed analysis using XP-EHH, and RsB Methods showed a greater number of highly selected markers and selected regions in Ladakhi cattle followed by Sahiwal, HF and Jersey (Table 4.19). The smaller number of common marker and selection signatures might be due to species difference, possibly. The Ladakhi cattle share more common markers and selection signatures with Ladakhi yak compared to Sahiwal, HF and Jersey, which might be due to the fact that they share the same climatic conditions and may share some evolutionary connection.

Table 4.19: Proportion of selected markers and shared selected regions between Ladakhi yak and cattle breeds

	Ladakhi yak vs Ladakhi cattle		**Ladakhi yak vs Sahiwal**		**Ladakhi yak vs HF**		**Ladakhi yak vs Jersey**	
Method	**SSM (p <0.05)**	**SSR**	**SSM (p <0.05)**	**SSR**	**SSM (p <0.05)**	**SSR**	**SSM (p <0.05)**	**SSR**
XP-EHH	297	34	145	31	140	32	21	7
RsB	308	37	128	35	137	38	19	6
XP-EHH+RsB	116	27	60	17	38	12	12	4

(SSM-Significant selected markers; SSR-Shared selected regions)

4.10.1 QTL mapping of shared selected regions between Ladakhi yak and cattle breeds

Shared selected regions between Ladakhi yak and cattle breeds were further mapped for QTLs and further classified trait wise. Among 27 Shared selected regions between Ladakhi yak and Ladakhi cattle, 21, 16 and 15 were mapped with QTLs associated with milk, reproduction, and adaptation traits (Table 4.20). In comparison with Sahiwal (17), as the number of shared regions QTL mapped with milk, reproduction and adaptation were 14, 12 and 11 (Table 4.21). In HF (12), the number of mapped regions for milk (10), reproduction (8), and adaptation (5) (Table 4.22) were lower than Ladakhi cattle. Similarly, Jersey (4) - was also having overall minimum shared selected regions associated with milk (4), reproduction (3) and adaptation (2) related traits (Table 4.23).

In this comparison we could not conclude highly positively selected shared regions between Ladakhi yak and cattle breeds due to few shared selected regions, hence we used common shared selected regions from both methods. Shared selected enumeration based on trait and could not give any detail difference due to its less number in the pair comparison between Ladakhi yak and cattle breeds. From the findings, it can be concluded that the vital traits such as cold tolerance and hematological parameters (Blood size and Blood volume) (Table 4.20) were found only in shared regions of Ladakhi cattle and Ladakhi yak. The finding is also in congruence with the similarity in physiological adaptation between these two species. The Ladakhi cattle is reared in the same climatic stress condition evolved in similar manner as such of Ladkahi yak and was able to acclimatize in same altitude as of yak. Relating to body conformation and external morphology, as our earlier observation indicates that the Ladakhi and Sahiwal cattle (Table 4.20; Table 4.21) has greater number of shared selected regions for these traits with Ladakhi yak, Ladakhi and Sahiwal cattle might develop similar external morphology on exposure to high altitude environment. Finally, the pairwise comparison of Ladakhi yak and exotic cattle breeds (Table 4.22; Table 4.23) had fewer selected regions and could not find any detailed findings to relate these.

Table 4.20: **Shared selected regions for milk, reproduction and adaptation traits between Ladakhi cattle and Ladakhi yak**

Chr.	Start	End	No. of Selected markers	Milk	Reproduction	Adaptation
7	6300000	6400000	2	Milk protein % Milk fat %	Calving ease, Scrotal circumference	Residual feed intake, Tick resistance, Cold tolerance
14	21700000	21800000	3	Milk protein yield	Calving ease	Residual feed intake
17	6500000	6600000	4	Milk protein %	Calving ease	PCV blood hemoglobin level
20	40100000	40200000	4	Milk fat %, Milk protein %	Calf size, Male fertility	Ear length, Muscularity
21	14900000	15000000	2	Milk protein %	Scrotal circumference	Structural soundness
23	25500000	25600000	3	Milk protein yield, Milk fat yield	Male fertility	Residual feed intake, Tick resistance
26	41200000	41300000	6	Milk protein %	Calving ease	Immunoglobulin level
27	37900000	38000000	6	Milk fat %, Milk protein %	Calving ease	Structural soundness
28	15400000	15500000	2	Milk yield	Calving ease, Scrotal circumference	Structural soundness

Table 4.21: **Shared Selected regions for milk, reproduction, and adaptation traits between Ladakhi yak and Sahiwal cattle**

Chr.	Start	End	No. of Selected markers	Milk	Reproduction	Adaptation
1	89700000	89800000	5	Milk fat yield, Milk composition	Maternal fertility	
2	62000000	62100000	5	Milk fat yield, Milk composition		Structural soundness
2	105500000	105600000	3	Milk protein %	Male fertility	
3	75800000	75900000	2		Calving ease	Structural soundness
5	25400000	25500000	2	Milk protein and fat %, Milk protein yield		Structural soundness
5	99600000	99700000	3	Milk fat yield, Milk composition		
8	17400000	17500000	5		Male fertility	Hematocrit (PCV %)
10	4300000	4400000	7	Milk composition		Body conformation
10	26800000	26900000	3	Milk protein %, Milk fat yield	Calving ease	Feet and leg conformation
14	21700000	21800000	2	Milk protein %, Milk protein yield	Calving ease	Tick Resistance, Residual Feed intake
14	46200000	46300000	5	Milk protein % and yield	Scrotal circumference	Tick resistance, Hematocrit (PCV %)
16	39900000	40000000	5		Calving ease	
17	14800000	14900000	3	Milk protein %	Calving ease	RBC number
17	33200000	33300000	2	Milk fat %	Calving ease	
19	12400000	12500000	4	Milk composition, Milk protein %		
23	28800000	29100000	4	Milk protein and fat yield	Calving ease	Tick resistance, Residual feed intake

Table 4.22: Shared selected regions for milk, reproduction and adaptation traits between Ladakhi yak and HF cattle

Chr.	Start	End	No. of Selected markers	Milk	Reproduction	Adaptation
3	43400000	43500000	2	Milk Protein %, Milk fat yield	Calving ease	Tick resistance
5	11000000	11100000	3	Milk fat %, Milk fat yield		Tick resistance
5	36700000	36800000	3	Milk fat %, Milk fat yield, Milk Protein %, Milk protein yield		Tick resistance
6	53500000	53600000	3	Milk fat yield, Milk fat %, Milk protein %	Maternal fertility, Calving ease	
10	83400000	83500000	3		Conception rate	
13	42300000	42400000	7	Milk protein yield		Residual Feed intake
14	7800000	7900000	2	Milk fat %, Milk protein %, Milk components	Scrotal circumference, Calving Ease	
15	20100000	20200000	2	Milk protein %		Structural soundness
16	27200000	27300000	2		Calving ease	
17	7100000	7200000	7	Milk protein%	Calving ease	Tick resistance
26	40100000	40200000	2	Milk protein yield	Calving ease, Calf size	

Table 4.23: Shared Selected regions for milk, reproduction, and adaptation traits between Ladakhi yak and Jersey cattle

Chr.	Start	End	No. of Selected markers	Milk	Reproduction	Adaptation
3	113500000	113600000	2	Milk protein %		
17	6500000	6600000	3	Milk protein %	Calving ease	Residual feed intake, PCV
23	25500000	25600000	5	Milk composition, Milk protein %	Scrotal circumference, Maternal Fertility	Residual feed intake, Tick resistance, Structural soundness
27	42900000	43000000	2	Milk composition	Calving ease	

CHAPTER –5

Summary and Conclusions

SUMMARY AND CONCLUSION

Ladakh (UT), a habitat with an altitude of more than 3000 m above mean sea level, is the home of unique farm bovine species, well adapted to hypoxic conditions, subnormal temperature, high UV exposure, and thriving on sparse feed resources. The unique genetic architects of these native bovines like cattle, yak and their crosses are an interesting venture to study the adaptation to high altitude region. Since, these bovines are low in milk productivity therefore, these also need to study about the genomic regions regulating the milk production, so that the regions can be exploited for possible genetic improvement. Genome wide comparison of these bovines with high producing zebu (Sahiwal) and exotic cattle (Holstein Friesian and Jersey) may be utilized for their genetic improvement for milk of native bovines without changing the genomic architect, conferring precious high altitude adaptation. The following objectives have been designed with the aim to explore the genetic architect and selection mechanism for milk and adaptation traits in high altitude bovines, and further comparison with high milk producing zebu and taurine breeds, using ddRAD sequence based *in silico* analysis:

1. To identify genetic variants and selection sweeps in Ladakhi cattle, yak and their hybrids
2. To functionally annotate genetic variants and selection sweeps for milk production and adaptation in high altitude bovines
3. To compare identified selection sweeps of high altitude adapted bovines with high producing *Bos taurus* and *Bos indicus* cattle

The study groups comprised- Ladakhi cattle (12), Ladakhi yak (12) and yak-cattle hybrids (6) of Ladakh (UT) region, and high producing cattle- Sahiwal (12), Holstein Friesian (12) and Jersey (12). To generate genomic raw data blood samples were collected, followed by isolation of genomic DNA, preparation of DNA library and finally ddRAD sequencing of the samples.

Raw data of the samples of these animals were checked for their quality using FastQC and Prinseq Lite 2.0 softwares. After initial quality check (QC) the Ladakhi cattle, Ladakhi yak and Yak cattle hybrid were found to have a total of 39.2, 29.5 and 82 11.78 million raw reads, respectively. The QC passed raw reads were demultiplexed and

allowed to remove low-quality reads by using Stacks 2.2. High quality raw reads were mapped with *Bos indicus* reference genome using Bowtie 2.0. The overall alignment rates with the reference genome were - 95.1 % in Ladakhi cattle, 86.2 % in Ladakhi yak and 90.3 % in Yak cattle hybrid, respectively. A variety of file format conversions were carried out using SAMtools, BAMtools, BCFtools and VCFtools to create variant calling file format (vcf file) for further analysis. The final file format was used to identify genome wide genetic variants such SNPs and InDels as well as downstream analysis. Genome wide variant calling at read depth (RD) 10 revealed 610058, 357981 and 311100 SNPs in Ladakhi cattle, Ladakhi yak and Yak hybrid, respectively. For screening of novel SNPs from prior reported SNPs *dbSNP 150 build* was used, which revealed 86.9 % novel SNPs in Ladakhi yak followed by 81.3 % in Yak-cattle hybrid and 59.9 % in Ladakhi cattle from the total identified SNPs for respective species.

The distribution and effect of identified genetic variants were explored through annotation using *SNPeff* tools. In analysis, the majority of variants were found to be substituted as transition (about 70 percent), with GA and CT predominant genotypes in all three bovines. Most of the variants were distributed in intronic (55 percent), followed by intergenic regions (40 percent) in all three bovines. Exonic coding regions accounted for about 1 percent of total identified variants. Maximum number of nucleotide changes were found to be synonymous (55 percent) type, followed by mis-sense (35 percent) or non-synonymous mutation causing the amino acid change. A total of 2848, 1087 and 857 mis-sense mutations were filtered out in Ladakhi cattle, Ladakhi yak and yak-cattle hybrid, respectively. These non-synonymous mutation were further mapped through gene annotation and found to located on 894, 410 and 362 genes in these three high altitude bovines, respectively.

Phenotype ontology analysis revealed that these mis-sense mutations were related with altering the structural (skeletal, cutaneous) and physiological (cardiovascular, respiratory, and nervous) changes, which corroborated with the high altitude adaptation of these bovines. A total of 56 genes mapped with mis-sense mutations were found common in all three high altitude bovines. Notably, polycystic kidney and hepatic disease 1 (PKHD1), BRCA1 interacting helicase 1 (BRIP1), galactosidase, beta 1 (GLB1), adenomatous polyposis coli (APC), fraser extracellular matrix complex subunit 1 (FRAS1), leucine rich repeat containing 6 (LRRC6), methylenetetrahydrofolate dehydrogenase cyclohydrolase and formyltetrahydrofolate synthetase 1 (MTHFD1),

NLR family pyrin domain containing 1 (NLRP1), radial spoke head component 4A (RSPH4A) and Tet methylcytosine dioxygenase 2 (TET2) genes were involved wih alteration of skeletal and skin coat texture related physiological functions. Whereas, Spectrin Repeat Containing Nuclear Envelope Protein 2 (SYNE2), collagen type IV alpha 3 chain (COL4A3) and WD repeat containing planar cell polarity effector (WDPCP) gene were found to be involved in physiological adaptation of respiratory and cardiovascular system to high altitude.

Genetic variants of these three high altitude bovines were further subjected to SNP density analysis using BAMtools. The results revealed that the Ladakhi cattle possessed the highest SNP density, with one SNP at every 4366 nucleotides, followed by the Ladakhi yak (one SNP at every 7488 nucleotides) and the yak cattle hybrid (one SNP at every 8563 nucleotides). The SNPs dense regions were calculated at 100 kb non overlapping windows. Total 253, 293 and 300 regions were found to have high SNP dense patterns in the Ladakhi cattle, yak and yak cattle hybrid, respectively. Among these, only 59 SNP rich regions were found to be common in these three bovines. The common SNPs were further mapped for the presence of *QTLs* associated with milk (39), reproduction (37) and adaptation (24) traits. Some of notable SNP dense regions were 70.0 - 70.39 Mb region of chromosome 12, rich with QTLs of milk composition; 68.4 - 68.49 Mb region of chromosome 7, rich with cold tolerance related QTLs, 1 - 1.0 Mb region of chromosome 1, related with conception rate and 43.9 - 43.99 Mb region of chromosome 17 related with milk composition and tick resistance in livestock.

Total 498426, 218470 and 267958 high quality SNPs were filtered out from total SNPs of Ladakhi cattle, Ladakhi yak and yak cattle hybrid, respectively after removing the variants with maximum missing genotype (<1), minor allele frequency (< 0.05) and Hardy Weinberg equilibrium (<0.0001), using VCFtools. These high-quality SNPs were further used to identify selection sweep in these species individually, using combined approach SweeD and iHS methods. A total of 4985, 1168 and 1945 selection sweeps were identified in Ladakhi cattle, Ladakhi yak and Yak cattle hybrid by iHS method, which were further annotated to 921, 342 and 423 genes, respectively. Among these, total 3825 selection sweeps were ancestral and 1160 as recent ones in Ladakhi cattle; whereas, 704 selection sweeps as ancestral and 465 as recent ones in Ladakhi yak. Using SweeD method, total 6346 selection signatures were identified in Ladakhi cattle, 2285 in Ladakhi yak and 2885 in Yak-cattle hybrid; annotated on 768, 544 and 523 genes,

respectively. Selection sweep scanning revealed 131, 78 and 67 genes with high significance in Ladakhi cattle, Ladakhi yak and Yak-cattle hybrid, respectively from both methods. Gene ontology and pathway analysis revealed that the high-altitude bovines developed a different set of mechanism to cope up with high-altitude conditions; for example, Ladakhi cattle primarily respond by PML, KCNMA1 and PRKCB genes; whereas Ladakhi yak responds by HIF1A, RYR2 and NOS2 genes to tackle the hypoxia. The co-selected genes among all three high altitude bovines were used to construct an adaptive interactive network for oxygen-carrying (PTPRM, CHRNA7, STARD13, LOXL2, and AMOTL1 gene), cardiovascular (PKP2, RYR2 and GJA1 gene), cutaneous (ADAMTS2, EGFR, SOS1, HUS1 and PTPRK gene) and energy metabolism system (IRS, AKT3, EGR1, PPARA, FTH and IGFBP5 gene), important for the cold hypoxic condition of high altitude.

Co-selected signatures were detected between high altitude bovines (Ladakhi cattle and Ladakhi yak) and high producing cattle (Sahiwal, Holstein Friesian and Jersey) using combinatorial approach of XP-EHH and RsB methods. The identified co-selected signatures were used to define the shared selected regions between high altitude bovines and high producing cattle through pair-wise comparison. These co-selected signatures were used to determine the genetic compatibility and similarities between the two groups based on the proportion of shared selected regions among them. During the pair-wise comparison of Ladakhi cattle with high producing cattle breeds, a total of 2055, 1645 and 1208 (co-)selected signatures were found to be common with the Sahiwal, HF and Jersey, respectively. Ladakhi cattle revealed total 539 positively selected regions common with Sahiwal, followed by HF (286) and Jersey (215). The Ladakhi cattle also had the highest number of common selected regions (with score more than 2), related to the milk, reproduction and adaptation traits shared with the Sahiwal (40, 25 and 28 regions, respectively) followed by the HF (23, 17 and 11 regions, respectively) and the Jersey (19, 15 and 17 regions, respectively). For adaptation, the shared vital traits between the Ladakhi and Sahiwal cattle were structural soundness (12 selected regions), fertility (11 shared selected regions) calving ease (7), residual feed intake (4) cold tolerance (3) and hematocrit (2). For milk production, the shared traits were: fat% and yield (17) protein % and yield (13) and milk yield (8) between these two cattle breeds. Pair-wise comparison of Ladakhi yak with cattle showed the highest number of common highly selected regions with the Ladakhi cattle (27), followed by the Sahiwal (17), HF

(12) and Jersey (4). The Ladakhi cattle found to share the greater proportion of QTLs with the Ladakhi yak in terms of milk (21), reproduction (16) and adaptation (15) traits compared to the Sahiwal, HF and Jersey.

Conclusion:

- The study indicated the common and converging genomic adaptation among high altitude bovines, probably due to the stringent selection process undergone to cope up with harsh conditions of high altitude.
- Greater proportion of novel mutations in Ladakhi cattle could suggest the genomic plasticity of the animals to changing climatic conditions.
- Hybridization of cattle and yak was found to affect the genomic composition in hybrids, however, these hybrids had more proximity with yak in terms of similar SNP dense patterns; associated with high altitude adaptation.
- Presence of individual Selection sweeps revealed that the high altitude bovines had developed common as well as unique set of mechanisms to adapt to high altitude conditions.
- Pattern of ancestral and recent selection sweep indicated about genome flexibility, which is still modulating to develop adaptive values, driven through environmental stressors in the high altitude bovines.
- Comparative analysis of high-altitude bovines with high producing cattle indicated a larger number of commonly selected regions between Ladakhi cattle and Sahiwal, providing a greater possibility for their better genomic compatibility than others, in case of crossbreeding.

Recommendations:

- This work highlights the genomic differences related with adaptive mechanism between Ladakhi cattle and yak. Further, a detailed study on this direction would help to sketch genetic architecture and evolutionary relations between the bovines adapted to high altitudes of Ladakh region.
- The role of newly reported genes- PTPRM, CHRNA7, STARD13, LOXL2 under this study, could be further examined for their role in high altitude climate adaptation including hypoxia.

- Strategy to compare the selection sweeps among bovines could be utilized for assessing the genomic compatibility, prior to any cross-breeding program for milk improvement.
- Presence of genomic regions conferring the high altitude adaptation may be considered to decide the best suited dairy cattle breeds at high altitude region of Ladakh.
- The milk production capability of Ladakhi cattle could be improved through opting crossbreeding, if opted, with most suitable high producing cattle like Sahiwal, based on the number of co-selection sweep.

Bibliography

BIBLIOGRAPHY

Aase, K., Ernkvist, M., Ebarasi, L., Jakobsson, L., Majumdar, A., Yi, C., ... and Holmgren, L. (2007). Angiomotin regulates endothelial cell migration during embryonic angiogenesis. Genes and development, 21(16), 2055-2068.

Ablondi, M., Viklund, Å., Lindgren, G., Eriksson, S., and Mikko, S. (2019). Signatures of selection in the genome of Swedish warmblood horses selected for sport performance. *BMC genomics*, *20*(1), 1-12.

Akand AH, Singh B and Chander M (2017). Contribution of livestock in livelihood: A case of Ladakh region. Indian Journal of Animal Sciences 87(5):649-652

Allen, W. T., Gower, S., and Wilsher, S. (2017). Localisation of epidermal growth factor (EGF), its specific receptor (EGF-R) and aromatase at the materno-fetal interface during placentation in the pregnant mare. Placenta, 50, 53-59.

Almeida, O. A. C., Moreira, G. C. M., Rezende, F. M., Boschiero, C., de Oliveira Peixoto, J., Ibelli, A. M. G., ... and Coutinho, L. L. (2019). Identification of selection signatures involved in performance traits in a paternal broiler line. *BMC genomics*, *20*(1), 1-20.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Avilés, C., Polvillo, O., Peña, F., Juárez, M., Martínez, A. L., and Molina, A. (2013). Associations between DGAT1, FABP4, LEP, RORC, and SCD1 gene polymorphisms in Israeli cattle for milk fat yield. Journal of animal science, 91(10), 4571-4577.

Ayalew, W., Chu, M., Liang, C., Wu, X., and Yan, P. (2021). Adaptation Mechanisms of Yak (Bos grunniens) to High-Altitude Environmental Stress. Animals, 11(8), 2344.

Bärtsch, P., and Gibbs, J. S. R. (2007). Effect of altitude on the heart and the lungs. Circulation, 116(19), 2191-2202.

Basak, P., Leslie, H., Dillon, R. L., Muller, W. J., Raouf, A., and Mowat, M. R. (2018). In vivo evidence supporting a metastasis suppressor role for Stard13 (Dlc2) in ErbB2 (Neu) oncogene induced mouse mammary tumors. Genes, Chromosomes and Cancer, 57(4), 182-191.

Basic Animal Husbandry Statistics (2019). Department of Animal Husbandry and Fisheries

Beall, C. M. (2007). Two routes to functional adaptation: Tibetan and Andean high-altitude natives. Proceedings of the National Academy of Sciences, 104(suppl 1), 8655-8660.

Ben-Jemaa, S., Mastrangelo, S., Lee, S. H., Lee, J. H., and Boussaha, M. (2020). Genome-wide scan for selection signatures reveals novel insights into the adaptive capacity in local North African cattle. Scientific reports, 10(1), 1-14.

Bharti, V. K., Giri, A., Vivek, P., and Kalia, S. (2017). Health and productivity of dairy cattle in high altitude cold desert environment of Leh-Ladakh: a review. Indian J Anim Sci, 87, 3-10.

Bigham, A. W., and Lee, F. S. (2014). Human high-altitude adaptation: forward genetics meets the HIF pathway. Genes and development, 28(20), 2189-2204.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., ... and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics, 25(8), 1091-1093. https://apps.cytoscape.org/apps/cluego

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., and SanCristobal, M. (2010). Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*, *186*(1), 241-262.

Bouverot P (2012). Adaptation to altitude-hypoxia in vertebrates (Vol. 16). Springer Science and Business Media.

Brito, L. F., Kijas, J. W., Ventura, R. V., Sargolzaei, M., Porto-Neto, L. R., Cánovas, A., ... and Schenkel, F. S. (2017). Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC genomics*, *18*(1), 1-20.

Bround, M. J., Asghari, P., Wambolt, R. B., Bohunek, L., Smits, C., Philit, M., ... and Johnson, J. D. (2012). Cardiac ryanodine receptors control heart rate and rhythmicity in adult mice. Cardiovascular research, 96(3), 372-380.

Cadzow, M., Boocock, J., Nguyen, H. T., Wilcox, P., Merriman, T. R., and Black, M. A. (2014). A bioinformatics workflow for detecting signatures of selection in genomic data. *Frontiers in Genetics, 5*, 293.

Cai, D., Choi, P. S., Gelbard, M., and Meyerson, M. (2019). Identification and characterization of oncogenic SOS1 mutations in lung adenocarcinoma. Molecular Cancer Research, 17(4), 1002-1012.

Canavez, F. C., Luche, D. D., Stothard, P., Leite, K. R., Sousa-Canavez, J. M., Plastow, G., ... and Camara-Lopes, L. H. (2012). Genome sequence and assembly of Bos indicus. Journal of Heredity, 103(3), 342-348.

Carlson, C. S., Thomas, D. J., Eberle, M. A., Swanson, J. E., Livingston, R. J., Rieder, M. J., and Nickerson, D. A. (2005). Genomic regions exhibiting positive selection identified from dense genotype data. *Genome research*, *15*(11), 1553-1565.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. Molecular ecology, 22(11), 3124-3140. http://catchenlab.life.illinois.edu/stacks/

Cavadas, M. A., Mesnieres, M., Crifo, B., Manresa, M. C., Selfridge, A. C., Keogh, C. E., ... and Cheong, A. (2016). REST is a hypoxia-responsive transcriptional repressor. Scientific reports, 6(1), 1-13.

Cerrone, M., Montnach, J., Lin, X., Zhao, Y. T., Zhang, M., Agullo-Pascual, E., ... and Delmar, M. (2017). Plakophilin-2 is required for transcription of genes that control calcium cycling and cardiac rhythm. Nature communications, 8(1), 1-16.

Cesarani, A., Sorbolini, S., Criscione, A., Bordonaro, S., Pulina, G., Battacone, G., ... and Macciotta, N. P. P. (2018). Genome- wide variability and selection signatures in Italian island cattle breeds. *Animal genetics*, *49*(5), 371-383.

Chang, L. S., Kim, M., Glinka, A., Reinhard, C., and Niehrs, C. (2020). The tumor suppressor PTPRK promotes ZNRF3 internalization and is required for Wnt inhibition in the Spemann organizer. Elife, 9, e51248.

Chen, M., Pan, D., Ren, H., Fu, J., Li, J., Su, G., ... and Liu, J. F. (2016). Identification of selective sweeps reveals divergent selection between Chinese Holstein and Simmental cattle populations. *Genetics Selection Evolution*, *48*(1), 1-12.

Chen, M., Wang, J., Wang, Y., Wu, Y., Fu, J., and Liu, J. F. (2018). Genome-wide detection of selection signatures in Chinese indigenous Laiwu pigs revealed candidate genes regulating fat deposition in muscle. *BMC genetics*, *19*(1), 1-9.

Cheruiyot, E. K., Bett, R. C., Amimo, J. O., Zhang, Y., Mrode, R., and Mujibi, F. D. (2018). Signatures of selection in admixed dairy cattle in Tanzania. Frontiers in genetics, 9, 607.

Chevin, L. M. (2019). Selective sweep at a QTL in a randomly fluctuating environment. Genetics, 213(3), 987-1005.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... and Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly, 6(2), 80-92.

Cirulli ET and Goldstein BD (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nature Reviews Genetics 11, no. 6: 415.

Cole, A. M., Cox, S., Jeong, C., Petousi, N., Aryal, D. R., Droma, Y., ... and Cavalleri, G. L. (2017). Genetic structure in the Sherpa and neighboring Nepalese populations. BMC genomics, 18(1), 1-10.

Colige, A., Sieron, A. L., Li, S. W., Schwarze, U., Petty, E., Wertelecki, W., ... and Nusgens, B. V. (1999). Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis are caused by mutations in the procollagen I N-proteinase gene. The American Journal of Human Genetics, 65(2), 308-317.

Curik, I., Ferenčaković, M., and Sölkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livestock Science*, *166*, 26-34.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... and Durbin, R. (2011). 1000 Genomes Project Analysis Group. 2011. The variant call

format and VCFtools. Bioinformatics, 27(15), 2156-2158. https://vcf tools.github.io/downloads.html

De Donato, M., Peters, S. O., Mitchell, S. E., Hussain, T., and Imumorin, I. G. (2013). Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. PloS one, 8(5), e62137.

Delseny M, Han B and Hsing Y (2010). High throughput DNA sequencing: the new sequencing revolution. Plant Science, 179(5): 407-422.

Department of Animal Husbandry and Fisheries, Government of India (2019). 20th livestock census report http://dahd.nic.in/

Diao, S., Luo, Y., Ma, Y., Deng, X., He, Y., Gao, N., Zhang, H., Li, J., Chen, Z and Zhang, Z., (2018). Genome-wide detection of selective signatures in a Duroc pig population. J. Integr. Agric. 17, 2528–2535.

Ding, X. Z., Liang, C. N., Guo, X., Wu, X. Y., Wang, H. B., Johnson, K. A., and Yan, P. (2014). Physiological insight into the high-altitude adaptations in domesticated yaks (Bos grunniens) along the Qinghai-Tibetan Plateau altitudinal gradient. Livestock Science, 162, 233-239.

Druet T, Pérez- Pardal L, Charlier C and Gautier M (2013). Identification of large selective sweeps associated with major genes in cattle. Animal Genetics, 44(6), 758-762.

Edea, Z., Dadi, H., Kim, S. W., Dessie, T., Lee, T., Kim, H., ... and Kim, K. S. (2013). Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. Frontiers in Genetics, 4, 35.

Edea, Z., Dadi, H., Kim, S. W., Park, J. H., Shin, G. H., Dessie, T., and Kim, K. S. (2014). Linkage disequilibrium and genomic scan to detect selective loci in cattle populations adapted to different ecological conditions in E thiopia. Journal of Animal Breeding and Genetics, 131(5), 358-366.

Egea, R., Casillas, S., and Barbadilla, A. (2008). Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic Acids Res 36, W157–W162.

Elshire RJ, Jeffrey CG, Qi S, Jesse AP, Ken K, Edward SB and Sharon EM (2013). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6, e19379.

Elsik CG, Deepak RU, Colin MD, Aditi T, Marianne LE, Hung NN and Darren E (2015). Bovine Genome Database: new tools for gleaning function from the Bos taurus genome. Nucleic Acids Research 44, D834-D839.

Elsik, C. G., Tayal, A., Diesh, C. M., Unni, D. R., Emery, M. L., Nguyen, H. N., and Hagen, D. E. (2016). Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine. Nucleic acids research, 44(D1), D793-D800.

Fariello, M.I., Boitard, S., Naya, H., SanCristobal, M., and Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics 193, 929–941.

Fariello, M.I., Servin, B., Tosser-Klopp, G., Rupp, R., Moreno, C., Cristobal, M.S., and Boitard, S. (2014). Selection signatures in worldwide sheep populations. PLoS One 9, e103813.

Fay, J.C., and Wu, C.I. (2000). Hitchhiking under positive Darwinian selection. Genetics 155, 1405–1413.

Friedrich, J., and Wiener, P. (2020). Selection signatures for high- altitude adaptation in ruminants. Animal genetics, 51(2), 157-165.

Gabriel, S., Ziaugra, L., and Tabbaa, D. (2009). SNP genotyping using the Sequenom MassARRAY iPLEX platform. Current protocols in human genetics, 60(1), 2-12.

Gall, L., Chene, N., Dahirel, M., Ruffini, S., and Boulesteix, C. (2004). Expression of epidermal growth factor receptor in the goat cumulus–oocyte complex. Molecular Reproduction and Development: Incorporating Gamete Research, 67(4), 439-445.

Gao, Y., Jiang, J., Yang, S., Hou, Y., Liu, G. E., Zhang, S., ... and Sun, D. (2017). CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. Bmc Genomics, 18(1), 1-12.

Gautier M and Naves M (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. Molecular Ecology 20, 3128–3143.

Gautier, M., and Vitalis, R. (2012). rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics, 28(8), 1176-1177. https://cran.r-project.org/web/packages/rehh/index.html

Ge, R. L., Simonson, T. S., Gordeuk, V., Prchal, J. T., and McClain, D. A. (2015). Metabolic aspects of high- altitude adaptation in Tibetans. Experimental physiology, 100(11), 1247-1255.

Ghoreishifar, S. M., Eriksson, S., Johansson, A. M., Khansefid, M., Moghaddaszadeh-Ahrabi, S., Parna, N., ... and Javanmard, A. (2020). Signatures of selection reveal candidate genes involved in economic traits and cold acclimation in five Swedish cattle breeds. Genetics Selection Evolution, 52(1), 1-15.

Gianola, D., Simianer, H and Qanbari, S., (2010). A two-step method for detecting selection signatures using genetic markers. Genet. Res. 92, 141–155.

Gibson, J., Morton, N.E and Collins, A (2006). Extended tracts of homozygosity in outbred human populations. Hum. Mol. Gen. 15, 789–795.

Goldman, N and Yang, Z. (1994). A codon-based model of nucleotide substitution for protein- coding DNA sequences. Mol. Biol. Evol. 11, 725–736.

Gonzalez-Candia, A., and Herrera, E. A. (2021). High altitude pregnancies and vascular dysfunction: observations from Latin American studies. Frontiers in physiology, 2124.

Gorkhali, N. A., Dong, K., Yang, M., Song, S., Kader, A., Shrestha, B. S., ... and Ma, Y. (2016). Genomic analysis identified a potential novel molecular mechanism for high-altitude adaptation in sheep at the Himalayas. Scientific Reports, 6(1), 1-10.

Graham, A. M., and McCracken, K. G. (2019). Convergent evolution on the hypoxia-inducible factor (HIF) pathway genes EGLN1 and EPAS1 in high-altitude ducks. Heredity, 122(6), 819-832.

Grivet, D., Avia, K., Vaattovaara, A., Eckert, A.J., Neale, D.B., Savolainen, O., and González-Martínez, S.C. (2017). High rate of adaptive evolution in two widespread European pines. Mol. Ecol. 26857–6870.

Grossman SR, Shlyakhter I, Karlsson EK, Byrne EH, Morales S.,.....Frieden G. (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327: 883–6.

Guang-Xin, E., Yang, B. G., Zhu, Y. B., Duang, X. H., Basang, W. D., Luo, X. L., and An, T. W. (2020). Genome-wide selective sweep analysis of the high-altitude adaptability of yaks by using the copy number variant. 3 Biotech, 10, 1-6.

Gurgul, A., Jasielczuk, I., Ropka-Molik, K., Semik-Gurgul, E., Pawlina-Tyszko, K., Szmatoła, T., Szyndler-Nędza, M., Bugno-Poniewierska, M., Blicharski, T., Szulc, K., Skrzypczak, and E., Krupiński, J., (2018). A genome-wide detection of selection signatures in conserved and commercial pig breeds maintained in Poland. BMC Genet 19, 95.

Gutiérrez-Gil B, Arranz JJ and Wiener P (2015). An interpretive review of selective sweep studies in Bos taurus cattle populations: identification of unique and shared selection signals across breeds. Frontiers in Genetics, 6, 167.

Hamilton JP and Robin Buell C (2012). Advances in plant genome sequencing. The Plant Journal, 70(1): 177-190.

Han, X. T., Xie, A. Y., Bi, X. C., Liu, S. J., and Hu, L. H. (2002). Effects of high altitude and season on fasting heat production in the yak Bos grunniens or Poephagus grunniens. British Journal of Nutrition, 88(2), 189-197.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., van Vuuren, B. J., Matthee, C., ... and Couloux, A. (2012). Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. Comptes rendus biologies, 335(1), 32-50.

Hawk, J. L. M., Young, A. R., and Ferguson, J. (2010). Cutaneous photobiology. Rook's Textbook of Dermatology, 1, 1-24.

Hayez, A., Roegiers, E., Malaisse, J., Balau, B., Sterpin, C., Achouri, Y., ... and De Backer, O. (2016). TMEM45A is dispensable for epidermal morphogenesis, keratinization and barrier formation. Plos one, 11(1), e0147069.

He, J., Zhao, X., Laroche, A., Lu, Z. X., Liu, H., and Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in plant science, 5, 484.

Heeschen, C., Weis, M., Aicher, A., Dimmeler, S., and Cooke, J. P. (2002). A novel angiogenic pathway mediated by non-neuronal nicotinic acetylcholine receptors. The Journal of clinical investigation, 110(4), 527-536.

Hermisson, J., and Pennings, P.S., (2005). Soft Sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169, 2335–2352. https://doi. org/10.1534/genetics.104.036947.

Hermisson, J., and Pennings, P.S., (2017). Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. Methods Ecol. Evol. 8, 700–716.

Holden, J. E., Stone, C. K., Clark, C. M., Brown, W. D., Nickles, R. J., Stanley, C., and Hochachka, P. W. (1995). Enhanced cardiac metabolism of plasma glucose in high-altitude natives: adaptation against chronic hypoxia. Journal of applied physiology, 79(1), 222-228.

Holsinger, K.E., (2004). Analysis of Genetic diversity in geographically structured populations: A Bayesian perspective. Hereditas 130, 245–255.

Honda, T., and Inui, M. (2020). PDZRN3 protects against apoptosis in myoblasts by maintaining cyclin A2 expression. Scientific reports, 10(1), 1-12.

Horikawa, S., Ishii, Y., Hamashima, T., Yamamoto, S., Mori, H., Fujimori, T., ... and Sasahara, M. (2015). PDGFRα plays a crucial role in connective tissue remodeling. Scientific reports, 5(1), 1-14.

Howard NP, van de WE, Tillman J, Tong CB, Silverstein KA and Luby JJ (2014). Two QTL characterized for soft scald and soggy breakdown in apple (Malus domestica) through pedigree-based analysis of a large population of interconnected families. Tree Genetics and Genomes, 14(1), 2.

Hu, X. J., Yang, J., Xie, X. L., Lv, F. H., Cao, Y. H., Li, W. R., ... and Li, M. H. (2019). The genome landscape of Tibetan sheep reveals adaptive introgression from argali and the history of early human settlements on the Qinghai–Tibetan Plateau. Molecular biology and evolution, 36(2), 283-303.

Hu, Z. L., and Reecy, J. M. (2007). Animal QTLdb: beyond a repository. Mammalian Genome, 18(1), 1-4.

Huang, C. W., Lin, Y. T., Ding, S. T., Lo, L. L., Wang, P. H., Lin, E. C., ... and Lu, Y. W. (2015). Efficient SNP discovery by combining microarray and lab-on-a-chip data for animal breeding and selection. Microarrays, 4(4), 570-595.

Huang, T., Zhou, Y., Zhang, J., Cheng, A. S., Yu, J., To, K. F., and Kang, W. (2018). The physiological role of Motin family and its dysregulation in tumorigenesis. Journal of translational medicine, 16(1), 1-13.

Hudson, R.R., Slatkin, M., and Maddison, W.P. (1992). Estimation of levels of gene flow from DNA sequence data. Genetics 132, 583–589.

Hwang, H. S., Hasdemir, C., Laver, D., Mehra, D., Turhan, K., Faggioni, M., ... and Knollmann, B. C. (2011). Inhibition of cardiac Ca2+ release channels (RyR2) determines efficacy of class I antiarrhythmic drugs in catecholaminergic polymorphic ventricular tachycardia. Circulation: Arrhythmia and Electrophysiology, 4(2), 128-135.

Igoshin, A.V., Yurchenko, A.A., Belonogova, N.M., Petrovsky, D.V., Aitnazarov, R.B., Soloshenko, V.A., Yudin, N.S., and Larkin, D.M., (2019). Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations. BMC Genet 20, 26.

Iqbal, N., Liu, X., Yang, T., Huang, Z., Hanif, Q., Asif, M., Khan QM. and Mansoor, S. (2019). Genomic variants identified from whole-genome resequencing of indicine cattle breeds from Pakistan. PloS one, 14(4), e0215065.

Ishisaka, M., and Hara, H. (2014). The roles of diacylglycerol kinases in the central nervous system: review of genetic studies in mice. Journal of pharmacological sciences, 124(3), 336-343.

Jaffey, J. A., Bullock, G., Teplin, E., Guo, J., Villani, N. A., Mhlanga- Mutangadura, T., ... and Johnson, G. S. (2019). A homozygous ADAMTS 2 nonsense mutation in a Doberman Pinscher dog with Ehlers Danlos syndrome and extreme skin fragility. Animal genetics, 50(5), 543-545.

Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D. B., Pritchard, J. K., ... and Di Rienzo, A. (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. Nature communications, 5(1), 1-7.

Jia, C., Kong, X., Koltes, J. E., Gou, X., Yang, S., Yan, D., ... and Wei, Z. (2016). Gene co-expression network analysis unraveling transcriptional regulation of high-altitude adaptation of Tibetan pig. PLoS One, 11(12), e0168161.

Jost, M., Kari, C., and Rodeck, U. (2000). The EGF receptor–an essential regulator of multiple epidermal functions. European Journal of Dermatology, 10(7), 505-10.

Kadri, N.K., Guldbrandtsen, B., Lund, M.S., and Sahana, G. (2015). Genetic dissection of milk yield traits and mastitis resistance quantitative trait loci on chromosome 20 in dairy cattle. J. Dairy Sci. 98, 9015–9025.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research, 28(1), 27-30.

Karim, L., Takeda, H., Lin, L., Druet, T., Arias, J. A., Baurain, D., ... and Coppieters, W. (2011). Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. Nature genetics, 43(5), 405-413.

Katagiri, S., and Takahashi, Y. (2004). Changes in EGF concentrations during estrous cycle in bovine endometrium and their alterations in repeat breeder cows. Theriogenology, 62(1-2), 103-112.Gharib-Hamrouche, N., Chene, N., and Martal, J. (1995). Comparative expression of TGF-α and EGF genes in the ovine conceptus and uterine endometrium in the peri-implantation period. Reproduction Nutrition Development, 35(3), 291-303.

Kelly, K. R., Williamson, D. L., Fealy, C. E., Kriz, D. A., Krishnan, R. K., Huang, H., ... and Kirwan, J. P. (2010). Acute altitude-induced hypoxia suppresses plasma glucose and leptin in healthy humans. Metabolism, 59(2), 200-205.

Kern, A.D., and Hahn, M.W. (2018). The Neutral Theory in Light of Natural Selection. Mol. Biol. Evol. 35, 1366–1371.

Khatkar MS, Thomson PC, Tammen I and Raadsma HW (2004). Quantitative trait loci mapping in dairy cattle: review and meta-analysis. Genetics Selection Evolution, 36 (2), 163.

Khayatzadeh N, Mészáros G, Utsunomiya Y, Garcia JF, Schnyder U Gredl, (2016). Locus-specific ancestry to detect recent response to selection in admixed Swiss Fleckvieh cattle. Animal Genetics 47, 637–646.

Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Porto Neto, L.R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J. and Dalrymple, B., (2012). Genome-wide analysis

of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol 10, e1001258.

Kim, J. W., Tchernyshyov, I., Semenza, G. L., and Dang, C. V. (2006). HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. Cell metabolism, 3(3), 177-185.

Kim, J.Y., Jeong, S., Kim, K.H., Lim, W.-J., Lee, H.-Y., and Kim, N. (2019). Discovery of genomic characteristics and selection signatures in Korean indigenous goats Through Comparison of 10 Goat Breeds. Front. Genet. 10, 699.

Kim, S., and Misra, A. (2007). SNP genotyping: technologies and biomedical applications. Annu. Rev. Biomed. Eng., 9, 289-320.

Kimura, M. (1968). Evolutionary rate at the molecular level. Nature 217, 624–626.

Kliem, A., Tetens, F., Klonisch, T., Grealy, M., and Fischer, B. (1998). Epidermal growth factor receptor and ligands in elongating bovine blastocysts. Molecular Reproduction and Development: Incorporating Gamete Research, 51(4), 402-412.

Kohno, T., Otsuka, A., Girard, L., Sato, M., Iwakawa, R., Ogiwara, H., ... and Yokota, J. (2010). A catalog of genes homozygously deleted in human lung cancer and the candidacy of PTPRD as a tumor suppressor gene. Genes, Chromosomes and Cancer, 49(4), 342-352.

Koopaee, H. K., and Koshkoiyeh, A. E. (2014). SNPs Genotyping technologies and their applications in farm animals breeding programs. Brazilian Archives of Biology and Technology, 57, 87-95.

Koves, T. R., Ussher, J. R., Noland, R. C., Slentz, D., Mosedale, M., Ilkayeva, O., ... and Muoio, D. M. (2008). Mitochondrial overload and incomplete fatty acid oxidation contribute to skeletal muscle insulin resistance. Cell metabolism, 7(1), 45-56.

Ku, C. S., Loy, E. Y., Salim, A., Pawitan, Y., and Chia, K. S. (2010). The discovery of human genetic variations and their use as disease markers: past, present and future. Journal of human genetics, 55(7), 403-415.

Kumar, D. R., Devadasan, M. J., Surya, T., Vineeth, M. R., Choudhary, A., Sivalingam, J., Kataria, R.S., Niranjan, S.K., Tantia, M.S. and Verma, A. (2020). Genomic

diversity and selection sweeps identified in Indian swamp buffaloes reveals it's uniqueness with riverine buffaloes. Genomics, 112(3), 2385-2392.

Lambert, S. R., Harwood, C. A., Purdie, K. J., Gulati, A., Matin, R. N., Romanowska, M., ... and Proby, C. M. (2012). Metastatic cutaneous squamous cell carcinoma shows frequent deletion in the protein tyrosine phosphatase receptor Type D gene. International journal of cancer, 131(3), E216-E226.

Lan, D., Xiong, X., Mipam, T. D., Fu, C., Li, Q., Ai, Y., ... and Li, J. (2018). Genetic diversity, molecular phylogeny, and selection evidence of Jinchuan yak revealed by whole-genome resequencing. G3: Genes, Genomes, Genetics, 8(3), 945-952.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology, 10(3), 1-10. http://bowtie bio.sourceforge.net/bowtie2.

Lefebvre, P., Chinetti, G., Fruchart, J. C., and Staels, B. (2006). Sorting out the roles of PPARα in energy metabolism and vascular homeostasis. The Journal of clinical investigation, 116(3), 571-580.

Lenfant C and Sullivan K (1971). Adaptation to high altitude. New England Journal of Medicine, 284(23), 1298-1309.

Lewontin, R.C., and Krakauer, J., (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74, 175–195.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, 27(21), 2987-2993. https://github.com/samtools/bcftools/

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics, 25(16), 2078-2079. https://github.com/samtools/samtools/

Li, J., Meng, X., Wang, L., Yu, Y., Yu, H., and Wei, Q. (2021). Changes in the expression levels of elastic fibres in yak lungs at different growth stages. BMC Developmental Biology, 21(1), 1-11.

Li, Z., Wang, Y., Zhang, M., Xu, P., Huang, H., Wu, D., and Meng, A. (2012). The Amotl2 gene inhibits Wnt/β-catenin signaling and regulates embryonic

development in zebrafish. Journal of Biological Chemistry, 287(16), 13005-13015.

Lindsay, B.G., (1988). Composite likelihood methods. Contemp. Math 80, 221–239.

Liu S, Vijayendran D and Bonning BC (2011). Next generation sequencing technologies for insect virus discovery for whole-genome analysis. Genomics 85(1):1-15.

Liu, J., Yuan, C., Guo, T., Wang, F., Zeng, Y., Ding, X., ... and Yang, B. (2020). Genetic signatures of high-altitude adaptation and geographic distribution in Tibetan sheep. Scientific reports, 10(1), 1-13.

Liu, L., Lin, J., and He, H. (2019). Identification of potential crucial genes associated with the pathogenesis and prognosis of endometrial cancer. Frontiers in genetics, 10, 373.

Liu, Y. P., Chen, W. D., Li, W. N., and Zhang, M. (2019). Overexpression of FNDC1 relates to poor prognosis and its knockdown impairs cell invasion and migration in gastric cancer. Technology in cancer research and treatment, 18, 1533033819869928.

Lorenzo FR, Chad H, Mikko M, Benjamin O, Sabina S, Tsewang T, Victor G et al., (2014). A genetic mechanism for Tibetan high-altitude adaptation. Nature Genetics 46, 951

Lu, N., Li, X., Tan, R., An, J., Cai, Z., Hu, X., ... and Lu, H. (2018). HIF-1α/Beclin1-mediated autophagy is involved in neuroprotection induced by hypoxic preconditioning. Journal of Molecular Neuroscience, 66(2), 238-250.

Lu, X. C. M., Williams, A. J., Yao, C., Berti, R., Hartings, J. A., Whipple, R., ... and Dave, J. R. (2004). Microarray analysis of acute and delayed gene expression profile in rats after focal ischemic brain injury and reperfusion. Journal of neuroscience research, 77(6), 843-857.

Lu, X., Zhao, Y., Wu, Z., Wen, T., Sun, L., and Ren, Z. (2019). Study on blood physiological characteristics of plateau adaptation in Tibetan sawa sheep. Southwest China Journal of Agricultural Sciences, 32(6), 1443-1447.

Luca, F., Hudson, R. R., Witonsky, D. B., and Di Rienzo, A. (2011). A reduced representation approach to population genetic analyses and applications to human evolution. Genome research, 21(7), 1087-1098.

Lühken, G. (2012). Genetic testing for phenotype-causing variants in sheep and goats. Molecular and cellular probes, 26(6), 231-237.

Lum, J. J., Bui, T., Gruber, M., Gordan, J. D., DeBerardinis, R. J., Covello, K. L., ... and Thompson, C. B. (2007). The transcription factor HIF-1α plays a critical role in the growth factor-dependent regulation of both aerobic and anaerobic glycolysis. Genes and development, 21(9), 1037-1049.

Ma, H., Zhang, S., Zhang, K., Zhan, H., Peng, X., Xie, S., Li, X., Zhao, S., and Ma, Y. (2019). Identifying selection signatures for backfat thickness in Yorkshire pigs highlights new regions affecting fat metabolism. Genes 10, 254. https://doi.org/10.3390/ genes10040254.

MacEachern, S., McEwan, J., McCulloch, A., Mather, A., Savin, K., and Goddard, M. (2009). Molecular evolution of the Bovini tribe (Bovidae, Bovinae): Is there evidence of rapid evolution or reduced selective constraint in Domestic cattle? BMC Genomics 10, 179. https://doi.org/10.1186/1471-2164-10-179.

Maiorano AM, Daniela LL, Shogo T, Alejandra MT, Nedenia BS, Yutaka M, Anibal E, Joslaine NS, Rogério AC and Josineudson AV (2018). Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. PLoS One 13, e0200694.

Majmundar, A. J., Wong, W. J., and Simon, M. C. (2010). Hypoxia-inducible factors and the response to hypoxic stress. Molecular cell, 40(2), 294-309.

Manou-Stathopoulou, V., Goodwin, C. D., Patterson, T., Redwood, S. R., Marber, M. S., and Williams, R. P. (2015). The effects of cold and exercise on the cardiovascular system. Heart, 101(10), 808-820.

Manzari, Z., Mehrabani- Yeganeh, H., Nejati- Javaremi, A., Moradi, M.H., and Gholizadeh, M., (2019). Detecting selection signatures in three Iranian sheep breeds. Anim. Genet. 50, 298–302. https://doi.org/10.1111/age.12772.

Mardis ER (2008). The impact of next-generation sequencing technology on genetics. Trends in genetics, 24(3): 133-14.

Martin, A., and Orgogozo, V. (2013). The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution, 67(5), 1235-1250.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P.,O'Connell, J., Moore, S.S., Smith, T.P.L., Sonstegard, T.S. and Van Tassell, C.P., (2009).Development and characterization of a high density SNP genotyping assay for cattle.PLoS One 4, e5350.

McDonald, J.H. and Kreitman, M., (1991). Adaptive protein evolution at the Adh locus in Drosophila. Nature 351, 652–654.

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac- Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A.K., Farrington, S.M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S.H., Dunlop, M.G., Wright, A.F., Campbell, H. and Wilson, J.F., (2008). Runs of Homozygosity in European Populations. Am. J. Hum. Genet. 83, 359–372.

Metzker ML (2010). Sequencing technologies - the next generation. Nature Reviews Genetics, 11(1): 31-46.

Moore, L. G. (2003). Fetal growth restriction and maternal oxygen transport during high altitude pregnancy. High Altitude Medicine and Biology, 4(2), 141-156.

Moore, L. G. (2021). HYPOXIA AND REPRODUCTIVE HEALTH: Reproductive challenges at high altitude: fertility, pregnancy and neonatal well-being. Reproduction, 161(1), F81-F90.

Moore, L. G., Shriver, M., Bemis, L., Hickler, B., Wilson, M., Brutsaert, T., ... and Vargas, E. (2004). Maternal adaptation to high-altitude pregnancy: an experiment of nature—a review. Placenta, 25, S60-S71.

Moravčíková, N., Kasarda, R., Vostrý, L., Krupová, Z., Krupa, E., Lehocká, K., Olšanská, B., Trakovická, A., Nádaský, R., Židek, R., Belej, Ľ. and Golian, J.(2019). Analysis of selection signatures in the beef cattle genome. Czech J. Anim. Sci. 64, 491–503.

Muchadeyi, F. C., Ibeagha-Awemu, E. M., Javaremi, A. N., Reynoso, G. A. G., Mwacharo, J., Rothschild, M. F., and Sölkner, J. Editorial: Why Livestock

Genomics for Developing Countries Offers Opportunities for Success. Frontiers in Genetics, 11(626), 1-2.

Mukherjee, S., Cai, Z., Mukherjee, A., Longkumer, I., Mech, M., Vupru, K., ... and Sahana, G. (2019). Whole genome sequence and de novo assembly revealed genomic architecture of Indian Mithun (Bos frontalis). BMC genomics, 20(1), 1-12.

Müller, I., Rössler, O. G., Wittig, C., Menger, M. D., and Thiel, G. (2012). Critical role of Egr transcription factors in regulating insulin biosynthesis, blood glucose homeostasis, and islet size. Endocrinology, 153(7), 3040-3053.

Naeije, R. (2010). Physiological adaptation of the cardiovascular system to high altitude. Progress in cardiovascular diseases, 52(6), 456-466.

Nam, S. Y., Ko, Y. S., Jung, J., Yoon, J., Kim, Y. H., Choi, Y. J., ... and Lee, B. L. (2011). A hypoxia-dependent upregulation of hypoxia-inducible factor-1 by nuclear factor-κ B promotes gastric tumour growth and angiogenesis. British journal of cancer, 104(1), 166-174.

Natarajan, C., Projecto-Garcia, J., Moriyama, H., Weber, R. E., Muñoz-Fuentes, V., Green, A. J., ... and Storz, J. F. (2015). Convergent evolution of hemoglobin function in high-altitude Andean waterfowl involves limited parallelism at the molecular sequence level. PLoS genetics, 11(12), e1005681.

Netzer, N., Strohl, K., Faulhaber, M., Gatterer, H., and Burtscher, M. (2013). Hypoxia- related altitude illnesses. Journal of travel medicine, 20(4), 247-255.

Nicholson, G., Smith, A.V., Jonsson, F., Gustafsson, O., Stefansson, K., and Donnelly, P., (2002). Assessing population differentiation and isolation from single-nucleotide polymorphism data. J. Royal Stat. Soc. Series B 64, 695–715.

Nicolazzi, E. L., Biffani, S., Biscarini, F., Orozco ter Wengel, P., Caprera, A., Nazzicari, N., and Stella, A. (2015). Software solutions for the livestock genomics SNP array revolution. Animal genetics, 46(4), 343-353.

Nielsen, R., (2005). Molecular signatures of natural selection. Annu. Rev. Genet. 39, 197–218.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., and Bustamante, C. (2005) b. Genomic scans for selective sweeps using SNP data. Genome research, 15(11), 1566-1575.

Nolte, W., Thaller, G., Kuehn, C., (2019). Selection signatures in four German warmblood horse breeds: Tracing breeding history in the modern sport horse. PLoS One 14, e0215913.

O'Brien, K. A., Simonson, T. S., and Murray, A. J. (2020). Metabolic adaptation to high altitude. Current Opinion in Endocrine and Metabolic Research, 11, 33-41

Ogorevc, J., Kunej, T., Razpet, A., and Dovc, P. (2009). Database of cattle candidate genes and genetic markers for milk production and mastitis. Animal genetics, 40(6), 832-851.

Oleksyk, T.K., Smith, M.W., and O'Brien, S.J., (2010.) Genome-wide scans for footprints of natural selection. Philos. Trans. R. Soc. B 365, 185–205.

Onzima, R.B., Upadhyay, M.R., Doekes, H.P., Brito, L.F., Bosse, M.,Kanis, E., Groenen, M.A.M., and Crooijmans, R.P.M.A., (2018). Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan goat breeds. Front. Genet. 9, 318.

Pan, Y., Cui, Y., Yu, S., Zhang, Q., Fan, J., Abdul Rasheed, B., and Yang, K. (2014). The Expression of Epidermal Growth Factor (EGF) and its Receptor (EGFR) During Post- Natal Testes Development in the Yak. Reproduction in Domestic Animals, 49(6), 970-976.

Papandreou, I., Cairns, R. A., Fontana, L., Lim, A. L., and Denko, N. C. (2006). HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption. Cell metabolism, 3(3), 187-197.

Paramio, M. T., and Izquierdo, D. (2014). Current status of in vitro embryo production in sheep and goats. Reproduction in Domestic Animals, 49, 37-48.

Parraguez, V. H., Urquieta, B., Pérez, L., Castellaro, G., De los Reyes, M., Torres-Rovira, L., ... and González-Bulnes, A. (2013). Fertility in a high-altitude environment is compromised by luteal dysfunction: the relative roles of hypoxia and oxidative stress. Reproductive Biology and Endocrinology, 11(1), 1-12.

Pastore, S., Mascia, F., Mariani, V., and Girolomoni, G. (2008). The epidermal growth factor receptor system in skin repair and inflammation. Journal of Investigative Dermatology, 128(6), 1365-1374.

Pavlidis, P., Živković, D., Stamatakis, A., and Alachiotis, N. (2013). SweeD: likelihood-based detection of selective sweeps in thousands of genomes. Molecular biology and evolution, 30(9), 2224-2234.

Pennings, P.S., Hermisson, J., (2006). Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. Mol. Biol. Evol. 23, 1076–1084.

Pérez-Pardal, L., Saglam, I. K., Costa, V., Chen, S., Miller, M. R., and Beja-Pereira, A. (2016). Comparison of high-density SNP chip versus RAD sequencing in cattle and related species. Journal of Animal Science, 94, 23-23.

Pescador, N., Villar, D., Cifuentes, D., Garcia-Rocha, M., Ortiz-Barahona, A., Vazquez, S., ... and del Peso, L. (2010). Hypoxia promotes glycogen accumulation through hypoxia inducible factor (HIF)-mediated induction of glycogen synthase 1. PloS one, 5(3), e9644.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS one, 7(5), e37135.

Pons- Guiraud, A. (2007). Dry skin in dermatology: a complex physiopathology. Journal of the European Academy of Dermatology and Venereology, 21, 1-4.

Pritchard, J.K., Pickrell, J.K., and Coop, G., (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr. Biol. 20, R208–R215. https:// doi.org/10.1016/j.cub.2009.11.055.

Puglisi, R., Cambuli, C., Capoferri, R., Giannino, L., Lukaj, A., Duchi, R., ... and Bongioni, G. (2013). Differential gene expression in cumulus oocyte complexes collected by ovum pick up from repeat breeder and normally fertile Holstein Friesian heifers. Animal reproduction science, 141(1-2), 26-33.

Purfield, D.C., McParland, S., Wall, E., and Berry, D.P., (2017). The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLoS One 12, e0176780.

Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, Moore S et al., (2011). Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. BMC Genomics,12:318.

Qanbari, S., Simianer, H., 2014. Mapping signatures of positive selection in the genome of livestock. Livest. Sci. 166, 133–143.

Qin, W., Liang, C. N., Guo, X., Chu, M., Pei, J., Bao, P. J., ... and Yan, P. (2015). PPARα signal pathway gene expression is associated with fatty acid content in yak and cattle longissimus dorsi muscle. Genet Mol Res, 14(4), 14469-78.

Qu, Y., Tian, S., Han, N., Zhao, H., Gao, B., Fu, J., ... and Lei, F. (2015). Genetic responses to seasonal variation in altitudinal stress: whole-genome resequencing of great tit in eastern Himalayas. Scientific Reports, 5(1), 1-10.

Randhawa, I. A., Khatkar, M. S., Thomson, P. C., and Raadsma, H. W. (2016). A meta-assembly of selection signatures in cattle. PloS one, 11(4), e0153013.

Rankin, E. B., Rha, J., Selak, M. A., Unger, T. L., Keith, B., Liu, Q., and Haase, V. H. (2009). Hypoxia-inducible factor 2 regulates hepatic lipid metabolism. Molecular and cellular biology, 29(16), 4527-4538.

Rebelato, A.B., and Caetano, A.R., (2018). Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. Pesqui. Agropecu. Bras. 53, 975–984.

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). g: Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic acids research, 35(suppl_2), W193-W200.https://biit.cs.ut.ee/gprofiler/gost

Rexroad, C., Vallet, J., Matukumalli, L. K., Reecy, J., Bickhart, D., Blackburn, H., ... and Wells, K. (2019). Genome to phenome: improving animal health, production, and well-being–a new USDA blueprint for animal genome research 2018–2027. Frontiers in genetics, 10, 327.

Reynolds, J., Weir, B.S., and Cockerham, C.C., (1983). Estimation of the coancestry coefficient:basis for a short-term genetic distance. Genetics 105, 767–779.

Roh, S. W., Abell, G. C., Kim, K. H., Nam, Y. D., and Bae, J. W. (2010). Comparing microarrays and next-generation sequencing technologies for microbial ecology research. Trends in biotechnology, 28(6), 291-299.

Roland, M. C. P., Friis, C. M., Godang, K., Bollerslev, J., Haugen, G., and Henriksen, T. (2014). Maternal factors associated with fetal growth and birthweight are independent determinants of placental weight and exhibit differential effects by fetal sex. PloS one, 9(2), e87303.

Ronen, R., Udpa, N., Halperin, E., and Bafna, V., (2013). Learning natural selection from the site frequency spectrum. Genetics 195, 181–193. https://doi.org/10.1534/genetics.113. 152587.

Rubin, C.J., Zody, M.C., Eriksson, J., Meadows, J.R.S., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallböök, F., Besnier, F., Carlborg, Ö.,Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K., and Andersson,L.,(2010). Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464, 587–591.

Russell, D.W. and Sambrook, J. (2001). Molecular cloning: a laboratory manual (Vol. 1, 112). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J. and Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. Nature, 419(6909), 832-837.

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., ... and Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations. Nature, 449(7164), 913-918.

Saleem, A., and Rajput, S. (2020). Insights from the in silico structural, functional and phylogenetic characterization of canine lysyl oxidase protein. Journal of Genetic Engineering and Biotechnology, 18(1), 1-11.

Sanders SJ and Mason CE (2016). The Newly Emerging View of the Genome. Genomics, Circuits, and Pathways in Clinical Neuropsychiatry, pp. 3-26. Academic Press.

Saravanan, K. A., Panigrahi, M., Kumar, H., Bhushan, B., Dutt, T., and Mishra, B. P. (2020). Selection signatures in livestock genome: A review of concepts, approaches and applications. Livestock Science, 104257.

Sato, M., Jiao, Q., Honda, T., Kurotani, R., Toyota, E., Okumura, S., ... and Ishikawa, Y. (2009). Activator of G protein signaling 8 (AGS8) is required for hypoxia-induced apoptosis of cardiomyocytes: role of Gβγ and connexin 43 (CX43). Journal of Biological Chemistry, 284(45), 31431-31440.

Scheinfeldt, L. B., Soi, S., Thompson, S., Ranciaro, A., Woldemeskel, D., Beggs, W., ... and Tishkoff, S. A. (2012). Genetic adaptation to high altitude in the Ethiopian highlands. Genome biology, 13(1), 1-9.

Schietke, R., Warnecke, C., Wacker, I., Schödel, J., Mole, D. R., Campean, V., ... and Wiesener, M. S. (2010). The lysyl oxidases LOX and LOXL2 are necessary and sufficient to repress E-cadherin in hypoxia: insights into cellular transformation processes mediated by HIF-1. Journal of Biological Chemistry, 285(9), 6658-6669.

Schmidt-Kastner, R., van Os, J., Steinbusch, H. W., and Schmitz, C. (2006). Gene regulation by hypoxia and the neurodevelopmental origin of schizophrenia. Schizophrenia research, 84(2-3), 253-271.

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics, 27(6), 863-864.http://prinseq.sourceforge.net/

Schweizer, R. M., Velotta, J. P., Ivy, C. M., Jones, M. R., Muir, S. M., Bradburd, G. S., ... and Cheviron, Z. A. (2019). Physiological and genomic evidence that selection on the transcription factor Epas1 has altered cardiovascular function in high-altitude deer mice. PLoS genetics, 15(11), e1008420.

Scortegagna, M., Ding, K., Oktay, Y., Gaur, A., Thurmond, F., Yan, L. J., ... and Garcia, J. A. (2003). Multiple organ pathology, metabolic abnormalities and impaired homeostasis of reactive oxygen species in Epas1−/− mice. Nature genetics, 35(4), 331-340.

Semenza, G. L. (2009). Involvement of oxygen-sensing pathways in physiologic and pathologic erythropoiesis. Blood, The Journal of the American Society of Hematology, 114(10), 2015-2019.

Shapiro, J.A., Huang, W., Zhang, C., Hubisz, M.J., Lu, J., Turissini, D.A., Fang, S., Wang, H.-Y., Hudson, R.R., Nielsen, R., Chen, Z., and Wu, C.-I., (2007). Adaptive genic evolution in the Drosophila genomes. Proc. Natl. Acad. Sci. U.S.A. 104, 2271–2276.

Shen, N., Jiang, S., Lu, J. M., Yu, X., Lai, S. S., Zhang, J. Z., ... and Li, C. J. (2015). The constitutive activation of Egr-1/C/EBPa mediates the development of type 2 diabetes mellitus by enhancing hepatic gluconeogenesis. The American journal of pathology, 185(2), 513-523.

Shergojry, S. A., Akhoon, Z. A., Mubarak, T., and Namgyal, D. (2017). Socio-economic impact of livestock in tribal areas of Leh. Journal of Krishi Vigyan, 6(1), 187-190.

Signer-Hasler, H., Burren, A., Neuditschko, M., Frischknecht, M., Garrick, D., Stricker, C., Gredler, B., Bapst, B., and Flury, C., (2017). Population structure and genomic inbreeding in nine Swiss dairy cattle populations. Genet. Sel. Evol. 49, 83. https://doi.org/10. 1186/s12711-017-0358-6.

Simonsen, K.L., Churchill, G.A., and Aquadro, C.F., (1995). Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141, 413–429.

Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., ... and Ge, R. (2010). Genetic evidence for high-altitude adaptation in Tibet. Science, 329(5987), 72-75.

Singh, L. C. G. (2017). High altitude dermatology. Indian journal of dermatology, 62(1), 59.

Sivalingam, J., Vineeth, M. R., Surya, T., Singh, K., Dixit, S. P., Niranjan, S. K., Tantia, M.S., Gupta, I.D. and Ravikumar, D. (2020). Genomic divergence reveals unique populations among Indian Yaks. Scientific reports, 10(1), 1-7.

Smith, J. M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genetics Research, 23(1), 23-35.

Solomon, D. A., Kim, J. S., Cronin, J. C., Sibenaller, Z., Ryken, T., Rosenberg, S. A., ... and Waldman, T. (2008). Mutational inactivation of PTPRD in glioblastoma multiforme and malignant melanoma. Cancer research, 68(24), 10300-10306.

Song, S., Yao, N., Yang, M., Liu, X., Dong, K., Zhao, Q., ... and Jiang, L. (2016). Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (Capra hircus). BMC genomics, 17(1), 1-12.

Storz, J. F., Scott, G. R., and Cheviron, Z. A. (2010). Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. Journal of Experimental Biology, 213(24), 4125-4136.

Sun, P. H., Ye, L., Mason, M. D., and Jiang, W. G. (2012). Protein tyrosine phosphatase μ (PTP μ or PTPRM), a negative regulator of proliferation and invasion of breast cancer cells, is associated with disease prognosis. PLoS One, 7(11), e50183.

Sun, Y. B., Fu, T. T., Jin, J. Q., Murphy, R. W., Hillis, D. M., Zhang, Y. P., and Che, J. (2018). Species groups distributed across elevational gradients reveal convergent and continuous genetic adaptation to high elevations. Proceedings of the National Academy of Sciences, 115(45), E10634-E10641.

Surya, T., Vineeth, M. R., Sivalingam, J., Tantia, M. S., Dixit, S. P., Niranjan, S. K., and Gupta, I. D. (2019). Genomewide identification and annotation of SNPs in Bubalus bubalis. Genomics, 111(6), 1695-1698

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Takatsu, K., Kuse, M., Yoshioka, S., and Acosta, T. J. (2018). Expression of epidermal growth factor (EGF) and its receptor in bovine endometrium throughout the luteal phase: effects of EGF on prostaglandin production in endometrial cells. Animal Reproduction (AR), 12(2), 328-335.

Tal, R. (2012). The role of hypoxia and hypoxia-inducible factor-1alpha in preeclampsia pathogenesis. Biology of reproduction, 87(6), 134-1.

Tamada, H., Tsubutani, D., Kawate, N., Inaba, T., Matsuyama, S., Imakawa, K. A. Z. U. H. I. K. O., ... and Sawada, T. (2002). Detection of transforming growth factor-α and epidermal growth factor receptor mRNA and immunohistochemical

localization of their proteins in the ovine uterus during the early implantation period. The Histochemical Journal, 34(8), 383-390.

Tang, K., Thornton, K. R., and Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. PLoS biology, 5(7), e171.

Taylor, J. F., Whitacre, L. K., Hoff, J. L., Tizioto, P. C., Kim, J., Decker, J. E., and Schnabel, R. D. (2016). Lessons for livestock genomics from genome and transcriptome sequencing in cattle and other mammals. Genetics Selection Evolution, 48(1), 1-18.

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., ... and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome research, 13(9), 2129-2141.

Tian, H., Luo, J., Shi, H., Chen, X., Wu, J., Liang, Y., ... and Loor, J. J. (2020). Role of peroxisome proliferator-activated receptor-α on the synthesis of monounsaturated fatty acids in goat mammary epithelial cells. Journal of animal science, 98(3), skaa062

Tibary, A., Anouassi, A., and Khatir, H. (2005). Update on reproductive biotechnologies in small ruminants and camelids. Theriogenology, 64(3), 618-638.

Tran, Q. T., Kennedy, L. H., Leon Carrion, S., Bodreddigari, S., Goodwin, S. B., Sutter, C. H., and Sutter, T. R. (2012). EGFR regulation of epidermal barrier function. Physiological genomics, 44(8), 455-469.

Tyagi, S., Gupta, P., Saini, A. S., Kaushal, C., and Sharma, S. (2011). The peroxisome proliferator-activated receptor: A family of nuclear receptors role in various diseases. Journal of advanced pharmaceutical technology and research, 2(4), 236.

Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., ... and Schön, C. C. (2014). A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. BMC genomics, 15(1), 1-15.

Vitti, J.J., Grossman, S.R., and Sabeti, P.C., (2013). Detecting natural selection in genomic data.Annu. Rev. Genet. 47, 97–120. https://doi.org/10.1146/annurev-genet-111212.

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. PLoS biology, 4(3), e72.

Vy, H.M.T., Kim, Y., 2015. A composite-likelihood method for detecting incomplete selective sweep from population genomic data. Genetics 200, 633–649.

Wagner, K., Mincheva, A., Korn, B., Lichter, P., and Pöpperl, H. (2001). Pbx4, a new Pbx family member on mouse chromosome 8, is expressed during spermatogenesis. Mechanisms of development, 103(1-2), 127-13

Wang, H., Long, R., Liang, J. B., Guo, X., Ding, L., and Shang, Z. (2011). Comparison of nitrogen metabolism in yak (Bos grunniens) and indigenous cattle (Bos taurus) on the Qinghai-Tibetan Plateau. Asian-Australasian Journal of Animal Sciences, 24(6), 766-773.

Wang, Y., Li, Z., Xu, P., Huang, L., Tong, J., Huang, H., and Meng, A. (2011) b. Angiomotin-like2 gene (amotl2) is required for migration and proliferation of endothelial cells during angiogenesis. Journal of Biological Chemistry, 286(47), 41095-41104.

Wei, C., Wang, H., Liu, G., Zhao, F., Kijas, J. W., Ma, Y., ... and Du, L. (2016). Genome-wide analysis reveals adaptation to high altitudes in Tibetan sheep. Scientific reports, 6(1), 1-11.

Weigand, H., and Leese, F., (2018). Detecting signatures of positive selection in non-model species using genomic data. Zool. J. Linn. Soc. 184, 528–583. https://doi.org/10. 1093/zoolinnean/zly007.

Weir, B.S., and Cockerham, C.C., (1984). Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370.

Weir, B.S., and Hill, W.G., (2002). Estimating F-statistics. Annu. Rev. Genet 36, 721–750. https://doi.org/10.1146/annurev.genet.36.050802.093940.

White, M. F. (2014). IRS2 integrates insulin/IGF1 signalling with metabolism, neurodegeneration and longevity. Diabetes, Obesity and Metabolism, 16(S1), 4-15.

Wiener, G., Han, J., and Long, R. (2003). The yak. FAO Regional office for Asia and the Pacific.

Williamson, S.H., Hubisz, M.J., Clark, A.G., Payseur, B.A., Bustamante, C.D., and Nielsen, R., (2007). Localizing recent adaptive evolution in the human genome. PLoS Genet 3, e90. https://doi.org/10.1371/journal.pgen.0030090.

Willing, E. M., Dreyer, C., and Van Oosterhout, C. (2012). Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers.

Witt, K. E., and Huerta-Sánchez, E. (2019). Convergent evolution in human and domesticate adaptation to high-altitude environments. Philosophical Transactions of the Royal Society B, 374(1777), 20180235.

Wright, S., (1949). The genetical structure of populations. Ann. Eugen. 15, 323–354.

Wu, D. D., Ding, X. D., Wang, S., Wójcik, J. M., Zhang, Y. I., Tokarska, M., ... and Zhang, Y. P. (2018). Pervasive introgression facilitated domestication and adaptation in the Bos species complex. Nature ecology and evolution, 2(7), 1139-1145.

Wu, D. D., Yang, C. P., Wang, M. S., Dong, K. Z., Yan, D. W., Hao, Z. Q., ... and Zhang, Y. P. (2020). Convergent genomic signatures of high-altitude adaptation among domestic mammals. National Science Review, 7(6), 952-963.

Xiong, X., Fu, M., Lan, D., Li, J., Zi, X., and Zhong, J. (2015). Yak response to high-altitude hypoxic stress by altering mRNA expression and DNA methylation of hypoxia-inducible factors. Animal biotechnology, 26(3), 222-229.

Xu Z, Gou X, Wang Z, Li N, Qiu F, Wei Z, Yan D, Yang S, Jia J, Kong X and Lu S (2014). Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. Genome Research. 1;24(8):1308-15.

Yang, C. B., Kiser, P. J., Zheng, Y. T., Varoqueaux, F., and Mower, G. D. (2007). Bidirectional regulation of Munc13–3 protein expression by age and dark rearing during the critical period in mouse visual cortex. Neuroscience, 150(3), 603-608.

Yang, G. Q., Chen, Y. M., Wang, J. P., Guo, C., Zhao, L., Wang, X. Y., Guo, Y., Li, L., Li, D.Z. and Guo, Z. H. (2016). Development of a universal and simplified

ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. Plant methods, 12(1), 1-17.

Yang, J. I., Li, W. R., Lv, F. H., He, S. G., Tian, S. L., Peng, W. F., ... and Liu, M. J. (2016). Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. Molecular biology and evolution, 33(10), 2576-2592.

Yang, J., Jin, Z. B., Chen, J., Huang, X. F., Li, X. M., Liang, Y. B., ... and Qu, J. (2017). Genetic signatures of high-altitude adaptation in Tibetans. Proceedings of the National Academy of Sciences, 114(16), 4189-4194.

Yang, W., Kang, X., Yang, Q., Lin, Y., and Fang, M. (2013). Review on the development of genotyping methods for assessing farm animal diversity. Journal of animal science and biotechnology, 4(1), 1-6.

Yin, Y., Lan, J. H., Nguyen, D., Valenzuela, N., Takemura, P., Bolon, Y. T., ... and Zhang, Q. (2016). Application of high-throughput next-generation sequencing for HLA typing on buccal extracted DNA: results from over 10,000 donor recruitment samples. PloS one, 11(10), e0165810.

Young, A. L., Abaan, H. O., Zerbino, D., Mullikin, J. C., Birney, E., and Margulies, E. H. (2010). A new strategy for genome assembly using short sequence reads and reduced representation libraries. Genome research, 20(2), 249-256.

Yu, X., Shen, N., Zhang, M. L., Pan, F. Y., Wang, C., Jia, W. P., ... and Li, C. J. (2011). Egr- 1 decreases adipocyte insulin sensitivity by tilting PI3K/Akt and MAPK signal balance in mice. The EMBO journal, 30(18), 3754-3765.

Yurchenko, A.A., Daetwyler, H.D., Yudin, N., Schnabel, R.D., Vander Jagt, C.J., Soloshenko, V., Lhasaranov, B., Popov, R., Taylor, J.F., and Larkin, D.M., (2018). Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. Sci. Rep 8, 12984. https://doi.org/10. 1038/s41598-018-31304-w.

Zamudio, S. (2007). High-altitude hypoxia and preeclampsia. Frontiers in bioscience: a journal and virtual library, 12, 2967.

Zhang, H., Wang, S. Z., Wang, Z. P., Da, Y., Wang, N., Hu, X. X., ... and Li, H. (2012). A genome-wide scan of selective sweeps in two broiler chicken lines divergently selected for abdominal fat content. BMC genomics, 13(1), 1-16.

Zhang, J., Zhang, Y., Sun, T., Guo, F., Huang, S., Chandalia, M., ... and Fu, M. (2013). Dietary obesity-induced Egr-1 in adipocytes facilitates energy storage via suppression of FOXC2. Scientific reports, 3(1), 1-10.

Zhang, W., Tong, H., Zhang, Z., Shao, S., Liu, D., Li, S., and Yan, Y. (2018). Transcription factor EGR1 promotes differentiation of bovine skeletal muscle satellite cells by regulating MyoG gene expression. Journal of cellular physiology, 233(1), 350-362.

Zhang, S., Liu, W., Liu, X., Du, X., Zhang, K., Zhang, Y., ... and Liu, J. (2021). Structural variants selected during yak domestication inferred from long-read whole-genome sequencing. Molecular Biology and Evolution.

Zhao, F., McParland, S., Kearney, F., Du, L., Berry, D.P., (2015). Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genet. Sel. Evol. 47, 49.

Zhao, Z., Fu, Y. X., Hewett-Emmett, D., and Boerwinkle, E. (2003). Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. Gene, 312, 207-213.

Zhou X, Youlin X, Ren X, Yulin C, Li H, Shunmou H, Boshou L, Yong L, Liyin Y, and Huifang J (2014). Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). BMC Genomics 15, 351

Zhou, D., Udpa, N., Ronen, R., Stobdan, T., Liang, J., Appenzeller, O., Zhao, H.W., Yin, Y., Du, Y., Guo, L. and Haddad, G. G. (2013). Whole-genome sequencing uncovers the genetic basis of chronic mountain sickness in Andean highlanders. The American Journal of Human Genetics, 93(3), 452-462

Zhou, H., Hickford, J. G., and Fang, Q. (2012). A premature stop codon in the ADAMTS2 gene is likely to be responsible for dermatosparaxis in Dorper sheep. Animal genetics, 43(4), 471-473.

Zhou, X., Wu, X., Chu, M., Liang, C., Ding, X., Pei, J., ... and Yan, P. (2020). Validation of suitable reference genes for gene expression studies on yak testis development. Animals, 10(2), 182.

Zhou, Z. Q., Zhao, J. J., Chen, C. L., Liu, Y., Zeng, J. X., Wu, Z. R., ... and Xia, J. C. (2019). HUS1 checkpoint clamp component (HUS1) is a potential tumor suppressor in primary hepatocellular carcinoma. Molecular carcinogenesis, 58(1), 76-87.

Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., ... and Salzberg, S. L. (2009). A whole-genome assembly of the domestic cow, Bos taurus. Genome biology, 10(4), 1-10.

Zwane, A. A., Schnabel, R. D., Hoff, J., Choudhury, A., Makgahlela, M. L., Maiwashe, A., ... and Taylor, J. F. (2019). Genome-wide SNP discovery in indigenous cattle breeds of South Africa. Frontiers in genetics, 10, 273.

Annexure

ANNEXURE

Preparation of Reagents

Tris (1 M) pH 8.0

Dissolve 121.10 gm of Tris base in 800 ml of double distilled waterand adjust the pH to 8.0 by adding concentrated HCl (around 40 ml). Allow the solution to cool to room temperature before making final volume to 1 litre and store at 4 ^{0}C.

Sodium chloride (5 M)

Dissolve 29.40 gm of Sodium chloride in 80 ml of double distilled water and make the final volume up to 100 ml. Sterilize by autoclaving and stored at 4 ^{0}C.

Sodium acetate (3 M)

Dissolve 24.60 gm of Sodium acetate (anhydrous) in 80 ml of double distilled water. Adjust the pH to 5.2 with glacial acetic acid and make up the final to 100 ml. Sterilize by autoclaving and store at room temperature.

EDTA (0.5 M)

Dissolve 18.60 gm of Disodium salt of EDTA in 80 ml of double distilled water by adding NaOH pellets. Adjust the pH to 8.0 and make up the final volume to 100 ml. Sterilize by autoclaving and store at 4 ^{0} C

Ammonium chloride (1 M)

Dissolve 5.35 gm ammonium chloride in 80 ml of double distilled water and make the final volume up to 100 ml. Sterilize by autoclaving and store at 4 0 C.

Potassium bicarbonate (1 M)

Dissolve 10.00 gm of sodium bicarbonate in 80 ml of double distilled water and make up to a final volume of 100 ml. Sterilize by autoclaving.

Sodium dodecyl sulphate (SDS) 10 %

Dissolve 10.00 gm of Sodium dodecyl sulphate in 80 ml of autoclaved double distilled water and make up to a final volume of 100 ml. No autoclaving needed.

Proteinase-K

Dissolve Proteinase-K (20 mg) in 1ml of double distilled water. Store at -20 ^{0}C

Ethanol 70 %

Ethanol 99.9 % : 70 ml

Distilled water : 30 ml

RBC Lysis buffer (1 X prepared freshly every time)

NH 4 Cl (155 mM) : 155.0 ml

KHCO 3n (10 mM) : 10.0 ml

EDTA (0.1 mM) : 2.0 ml

Make up to 1000 ml by double distilled water. Sterilize by autoclaving and store at 4 ^{0}C

Phenol equilibration (Tris saturation)

Liquify Phenol crystals (500 gm) stored at –20 ^{0}C by keeping in a water bath maintained at 65 ^{0}C for 1 hour. Add 8-hydroxyquinoline to liquefied phenol at a final concentration of 0.1 %. Add equal volume (500 ml) of 0.5 M Tris (pH 8.0) and stir for 4 hours on magnetic stirrer, check pH repeatedly till it reached 8.0. Finally, 0.1 M Tris was added to an equilibrated phenol and stirred well and stored in amber colored bottles at (4 ^{0}C).

TE buffer (10 mM)

Tris (1M, pH 8.0) : 1.00 ml

EDTA (0.5M, pH 8.0) : 200 µl

Final volume made up to : 100 ml

Sterilize by autoclaving and store the buffer at room temperature.

TAE buffer (50X)

Tris base : 242.0 gm

Glacial acetic acid : 57.1 ml

EDTA (0.5 M, pH 8.0) : 100 ml

Add double distilled water to make the final volume 1000ml, filter and autoclave.